GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:46:55 ; Search time 38.34 Seconds
(without alignments)
113.989 Million cell updates/sec

Title: US-09-710-239-18
Perfect score: 333
Sequence: 1 EAGLPGAKGLTGSPGSPGPD.....PPGARGQAGVMGFPGPKGAA 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	333	100.0	59	22	AAE02704	Human alphal (I) t
2	333	100.0	59	22	AAB68058	Amino acid sequenc
3	333	100.0	101	22	AAE02705	Human alphal (I) t
4	333	100.0	101	22	AAB68059	Amino acid sequenc
5	333	100.0	185	22	AAE02706	Human alphal (I) t
6	333	100.0	185	22	AAB68060	Amino acid sequenc
7	333	100.0	251	22	AAE02707	Human alphal (I) t
8	333	100.0	251	22	AAB68061	Amino acid sequenc
9	333	100.0	500	22	AAE02708	Human alphal (I) t
10	333	100.0	500	22	AAB68062	Amino acid sequenc
11	333	100.0	501	22	AAE02703	Human alphal (I) t
12	333	100.0	501	22	AAB68057	Amino acid sequenc
13	333	100.0	662	22	AAE02718	Human alphal (I) t
14	333	100.0	662	22	AAB68072	Amino acid sequenc
15	333	100.0	1057	21	AAY84541	Amino acid sequenc
16	333	100.0	1057	21	AAY84544	A human collagen 1
17	333	100.0	1058	21	AAY84403	Amino acid sequenc
18	333	100.0	1107	17	AAR89472	Collagen/decorin(a
19	333	100.0	1107	21	AAY84540	Amino acid sequenc
20	333	100.0	1169	17	AAR89469	Collagen/BMP-2B fu
21	333	100.0	1169	21	AAY84537	Amino acid sequenc
22	333	100.0	1171	17	AAR89470	Collagen/TGF-beta-
23	333	100.0	1171	21	AAY84538	A chimeric collage
24	333	100.0	1341	16	AAR71701	Collagen alpha 1 (
25	333	100.0	1341	21	AAY96122	Collagen type I al
26	333	100.0	1388	17	AAR89471	Collagen/decorin f
27	333	100.0	1411	21	AAY56800	Human preproalpha
28	333	100.0	1449	22	AAE02535	Porcine alpha1(I)
29	333	100.0	1463	22	AAE02532	Bovine alpha1(I) c
30	333	100.0	1464	19	AAW68485	Human recombinant
31	333	100.0	1464	22	AAU14136	Human novel protei
32	333	100.0	1464	22	AAB82454	Human pro-alpha-1
33	325	97.6	1388	21	AAY84539	Amino acid sequenc
34	321	96.4	595	20	AAY06239	Mouse recombinant
35	321	96.4	822	20	AAY06240	Mouse recombinant
36	256	76.9	1418	15	AAR59751	Type II collagen.
37	256	76.9	1418	16	AAR71703	Collagen alpha 1 (
38	256	76.9	1418	21	AAY96124	Collagen type II a
39	256	76.9	1418	22	AAB35624	Human type II coll
40	256	76.9	1442	16	AAR79480	Rat type II collag
41	256	76.9	1487	19	AAW61562	Human type II coll
42	234	70.3	1078	16	AAR71704	Collagen alpha 1 (
43	234	70.3	1078	21	AAY96125	Collagen type III
44	234	70.3	1196	13	AAR28916	Type III procollag
45	234	70.3	1466	22	AAE02537	Porcine alpha1(III

ALIGNMENTS

RESULT 1
AAE02704
ID AAE02704 standard; Protein; 59 AA.
XX
AC AAE02704;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human alphal (I) type I collagen helical domain (residues 531-589).
XX
KW Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW encapsulant; film-forming agent; moisturising agent; thickening agent;
KW gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW plasma expander; colloidal volume replacement material; graft coating;
KW medical sponge; medical plug; micro-carrier; edible composition;
KW protein supplement; fat substitute; nutritional supplement; cell culture;
KW edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW cartilage degeneration; joint flexibility; food industry; beverage;
KW alphal (I) type I collagen.
XX
OS Homo sapiens.
XX
PN WO200134646-A2.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-US30791.
XX
PR 12-NOV-1999; 99US-0165114.
PR 15-MAY-2000; 2000US-0204437.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;

XX
DR WPI; 2001-329072/34.
XX
PT Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT prepared recombinantly -
XX
PS Claim 21; Page 123; 137pp; English.
XX
CC The patent discloses recombinant human gelatin which is useful
CC in various compositions including binding agents, encapsulants,
CC stabilising agents, film-forming agents, moisturising agents,
CC emulsifiers, thickening agents, gelling agents, colloidal agents,
CC adhesive agents, pharmaceutical compositions, hard gel capsules,
CC soft gel capsules, plasma expander, colloidal volume replacement
CC materials, graft coatings, medical sponges, medical plugs,
CC pharmaceutical stabilisers, micro-carriers, edible compositions,
CC protein supplements, fat substitutes, nutritional supplements,
CC edible coatings, photographic compositions, cosmetic compositions,
CC industrial composition, cell culture compositions and compositions
CC for use in the laboratory. Pharmaceutical compositions comprising
CC recombinant gelatin are used as vaccines. They are also used to
CC treat various joint conditions such as arthritis, athrosis and
CC other conditions related to the degeneration of cartilage and joint
CC flexibility. Recombinant gelatin is also used in food and beverage
CC industries. The present sequence is human alpha1 (I) type I collagen
CC helical domain (residues 531-589). This sequence is a recombinant
CC gelatin.
XX
SQ Sequence 59 AA;

Query Match 100.0%; Score 333; DB 22; Length 59;
Best Local Similarity 100.0%; Pred. No. 5.6e-24;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgpargqagvmgfpgpkgaa 59

RESULT 2
AAB68058
ID AAB68058 standard; Protein; 59 AA.
XX
AC AAB68058;
XX
DT 09-JUL-2001 (first entry)
XX
DE Amino acid sequence of a recombinant human gelatin.
XX
KW Human; gelatin; vaccine; anaphylactic reaction.
XX
OS Homo sapiens.
XX
PN WO200134801-A2.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-US30843.
XX
PR 12-NOV-1999; 99US-0165114.
PR 15-MAY-2000; 2000US-0204437.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
XX
DR WPI; 2001-308784/32.
XX
PT Vaccine formulations (I) comprising recombinant human gelatin, useful
PT for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT and cholera, the gelatin is non-immunogenic and confers stability at
PT ambient temperatures -
XX
PS Claim 11; Page 116; 130pp; English.
XX
CC The present sequence represents a human recombinant gelatin polypeptide.
CC The recombinant gelatin polypeptide is used to produce vaccine
CC formulations of the invention. The recombinant human gelatin is
CC non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC stability at ambient temperatures. The vaccine formulation comprises a
CC vaccine formulated for the prevention of a disease selected from vaccinia
CC virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC haemophillus influenzae meningitis, rabies, cholera, Japanese
CC encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ Sequence 59 AA;

Query Match 100.0%; Score 333; DB 22; Length 59;
Best Local Similarity 100.0%; Pred. No. 5.6e-24;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgpargqagvmgfpgpkgaa 59

RESULT 3
AAE02705
ID AAE02705 standard; Protein; 101 AA.
XX
AC AAE02705;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human alpha1 (I) type I collagen helical domain (residues 531-631).
XX
KW Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW encapsulant; film-forming agent; moisturising agent; thickening agent;
KW gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW plasma expander; colloidal volume replacement material; graft coating;
KW medical sponge; medical plug; micro-carrier; edible composition;
KW protein supplement; fat substitute; nutritional supplement; cell culture;
KW edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW cartilage degeneration; joint flexibility; food industry; beverage;
KW alpha1 (I) type I collagen.
XX
OS Homo sapiens.
XX
PN WO200134646-A2.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-US30791.
XX
PR 12-NOV-1999; 99US-0165114.
PR 15-MAY-2000; 2000US-0204437.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
XX
DR WPI; 2001-329072/34.
XX
PT Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT prepared recombinantly -
XX
PS Claim 21; Page 123-124; 137pp; English.
XX
CC The patent discloses recombinant human gelatin which is useful

CC in various compositions including binding agents, encapsulants,
CC stabilising agents, film-forming agents, moisturising agents,
CC emulsifiers, thickening agents, gelling agents, colloidal agents,
CC adhesive agents, pharmaceutical compositions, hard gel capsules,
CC soft gel capsules, plasma expander, colloidal volume replacement
CC materials, graft coatings, medical sponges, medical plugs,
CC pharmaceutical stabilisers, micro-carriers, edible compositions,
CC protein supplements, fat substitutes, nutritional supplements,
CC edible coatings, photographic compositions, cosmetic compositions,
CC industrial composition, cell culture compositions and compositions
CC for use in the laboratory. Pharmaceutical compositions comprising
CC recombinant gelatin are used as vaccines. They are also used to
CC treat various joint conditions such as arthritis, athrosis and
CC other conditions related to the degeneration of cartilage and joint
CC flexibility. Recombinant gelatin is also used in food and beverage
CC industries. The present sequence is human alpha1 (I) type I collagen
CC helical domain (residues 531-631). This sequence is a recombinant
CC gelatin.
XX
SQ Sequence 101 AA;

Query Match 100.0%; Score 333; DB 22; Length 101;
Best Local Similarity 100.0%; Pred. No. 9.3e-24;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 59

RESULT 4
AAB68059
ID AAB68059 standard; Protein; 101 AA.
XX
AC AAB68059;
XX
DT 09-JUL-2001 (first entry)
XX
DE Amino acid sequence of a recombinant human gelatin.
XX
KW Human; gelatin; vaccine; anaphylactic reaction.
XX
OS Homo sapiens.
XX
PN WO200134801-A2.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-US30843.
XX
PR 12-NOV-1999; 99US-0165114.
PR 15-MAY-2000; 2000US-0204437.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
XX
DR WPI; 2001-308784/32.
XX
PT Vaccine formulations (I) comprising recombinant human gelatin, useful
PT for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT and cholera, the gelatin is non-immunogenic and confers stability at
PT ambient temperatures -
XX
PS Claim 11; Page 116-117; 130pp; English.
XX
CC The present sequence represents a human recombinant gelatin polypeptide.
CC The recombinant gelatin polypeptide is used to produce vaccine
CC formulations of the invention. The recombinant human gelatin is
CC non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC stability at ambient temperatures. The vaccine formulation comprises a
CC vaccine formulated for the prevention of a disease selected from vaccinia
CC virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC haemophillus influenzae meningitis, rabies, cholera, Japanese
CC encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ Sequence 101 AA;

Query Match 100.0%; Score 333; DB 22; Length 101;
Best Local Similarity 100.0%; Pred. No. 9.3e-24;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 59

RESULT 5
AAE02706
ID AAE02706 standard; Protein; 185 AA.
XX
AC AAE02706;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human alpha1 (I) type I collagen helical domain (residues 531-715).
XX
KW Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW encapsulant; film-forming agent; moisturising agent; thickening agent;
KW gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW plasma expander; colloidal volume replacement material; graft coating;
KW medical sponge; medical plug; micro-carrier; edible composition;
KW protein supplement; fat substitute; nutritional supplement; cell culture;
KW edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW cartilage degeneration; joint flexibility; food industry; beverage;
KW alpha1 (I) type I collagen.
XX
OS Homo sapiens.
XX
PN WO200134646-A2.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-US30791.
XX
PR 12-NOV-1999; 99US-0165114.
PR 15-MAY-2000; 2000US-0204437.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
XX
DR WPI; 2001-329072/34.
XX
PT Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT prepared recombinantly -
XX
PS Claim 21; Page 124; 137pp; English.
XX
CC The patent discloses recombinant human gelatin which is useful
CC in various compositions including binding agents, encapsulants,
CC stabilising agents, film-forming agents, moisturising agents,
CC emulsifiers, thickening agents, gelling agents, colloidal agents,
CC adhesive agents, pharmaceutical compositions, hard gel capsules,
CC soft gel capsules, plasma expander, colloidal volume replacement
CC materials, graft coatings, medical sponges, medical plugs,
CC pharmaceutical stabilisers, micro-carriers, edible compositions,
CC protein supplements, fat substitutes, nutritional supplements,
CC edible coatings, photographic compositions, cosmetic compositions,

```
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is human alphal (I) type I collagen
CC   helical domain (residues 531-715). This sequence is a recombinant
CC   gelatin.
XX
SQ   Sequence   185 AA;


  Query Match             100.0%;  Score 333;  DB 22;  Length 185;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 59


RESULT   6
AAB68060
ID   AAB68060 standard; Protein; 185 AA.
XX
AC   AAB68060;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 117; 130pp; English.
XX
CC   The present sequence represents a human recombinant gelatin polypeptide.
CC   The recombinant gelatin polypeptide is used to produce vaccine
CC   formulations of the invention. The recombinant human gelatin is
CC   non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC   stability at ambient temperatures. The vaccine formulation comprises a
CC   vaccine formulated for the prevention of a disease selected from vaccinia
CC   virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC   diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC   (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC   haemophillus influenzae meningitis, rabies, cholera, Japanese
CC   encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC   adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC   respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC   herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ   Sequence   185 AA;


  Query Match             100.0%;  Score 333;  DB 22;  Length 185;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 59


RESULT   7
AAE02707
ID   AAE02707 standard; Protein; 251 AA.
XX
AC   AAE02707;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Human alphal (I) type I collagen helical domain (residues 531-781).
XX
KW   Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW   encapsulant; film-forming agent; moisturising agent; thickening agent;
KW   gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW   plasma expander; colloidal volume replacement material; graft coating;
KW   medical sponge; medical plug; micro-carrier; edible composition;
KW   protein supplement; fat substitute; nutritional supplement; cell culture;
KW   edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW   cartilage degeneration; joint flexibility; food industry; beverage;
KW   alphal (I) type I collagen.
XX
OS   Homo sapiens.
XX
PN   WO200134646-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30791.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-329072/34.
XX
PT   Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT   prepared recombinantly -
XX
PS   Claim 21; Page 125; 137pp; English.
XX
CC   The patent discloses recombinant human gelatin which is useful
CC   in various compositions including binding agents, encapsulants,
CC   stabilising agents, film-forming agents, moisturising agents,
CC   emulsifiers, thickening agents, gelling agents, colloidal agents,
CC   adhesive agents, pharmaceutical compositions, hard gel capsules,
CC   soft gel capsules, plasma expander, colloidal volume replacement
CC   materials, graft coatings, medical sponges, medical plugs,
CC   pharmaceutical stabilisers, micro-carriers, edible compositions,
CC   protein supplements, fat substitutes, nutritional supplements,
CC   edible coatings, photographic compositions, cosmetic compositions,
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is human alphal (I) type I collagen
CC   helical domain (residues 531-781). This sequence is a recombinant
CC   gelatin.
```

```
XX
SQ   Sequence   251 AA;


  Query Match           100.0%;  Score 333;  DB 22;  Length 251;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 59


RESULT   8
AAB68061
ID   AAB68061 standard; Protein; 251 AA.
XX
AC   AAB68061;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 118; 130pp; English.
XX
CC   The present sequence represents a human recombinant gelatin polypeptide.
CC   The recombinant gelatin polypeptide is used to produce vaccine
CC   formulations of the invention. The recombinant human gelatin is
CC   non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC   stability at ambient temperatures. The vaccine formulation comprises a
CC   vaccine formulated for the prevention of a disease selected from vaccinia
CC   virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC   diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC   (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC   haemophillus influenzae meningitis, rabies, cholera, Japanese
CC   encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC   adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC   respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC   herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ   Sequence   251 AA;


  Query Match           100.0%;  Score 333;  DB 22;  Length 251;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 59


RESULT   9
AAE02708
ID   AAE02708 standard; Protein; 500 AA.
XX
AC   AAE02708;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Human alpha1 (I) type I collagen helical domain (residues 531-1030).
XX
KW   Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW   encapsulant; film-forming agent; moisturising agent; thickening agent;
KW   gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW   plasma expander; colloidal volume replacement material; graft coating;
KW   medical sponge; medical plug; micro-carrier; edible composition;
KW   protein supplement; fat substitute; nutritional supplement; cell culture;
KW   edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW   cartilage degeneration; joint flexibility; food industry; beverage;
KW   alpha1 (I) type I collagen.
XX
OS   Homo sapiens.
XX
PN   WO200134646-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30791.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-329072/34.
XX
PT   Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT   prepared recombinantly -
XX
PS   Claim 21; Page 125-127; 137pp; English.
XX
CC   The patent discloses recombinant human gelatin which is useful
CC   in various compositions including binding agents, encapsulants,
CC   stabilising agents, film-forming agents, moisturising agents,
CC   emulsifiers, thickening agents, gelling agents, colloidal agents,
CC   adhesive agents, pharmaceutical compositions, hard gel capsules,
CC   soft gel capsules, plasma expander, colloidal volume replacement
CC   materials, graft coatings, medical sponges, medical plugs,
CC   pharmaceutical stabilisers, micro-carriers, edible compositions,
CC   protein supplements, fat substitutes, nutritional supplements,
CC   edible coatings, photographic compositions, cosmetic compositions,
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is human alpha1 (I) type I collagen
CC   helical domain (residues 531-1030). This sequence is a recombinant
CC   gelatin.
XX
SQ   Sequence   500 AA;


  Query Match           100.0%;  Score 333;  DB 22;  Length 500;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 59
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgpargqagvmgfpgpkgaa 59


RESULT  10
AAB68062
ID   AAB68062 standard; Protein; 500 AA.
XX
AC   AAB68062;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 118-120; 130pp; English.
XX
CC   The present sequence represents a human recombinant gelatin polypeptide.
CC   The recombinant gelatin polypeptide is used to produce vaccine
CC   formulations of the invention. The recombinant human gelatin is
CC   non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC   stability at ambient temperatures. The vaccine formulation comprises a
CC   vaccine formulated for the prevention of a disease selected from vaccinia
CC   virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC   diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC   (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC   haemophillus influenzae meningitis, rabies, cholera, Japanese
CC   encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC   adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC   respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC   herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ   Sequence   500 AA;


  Query Match             100.0%;  Score 333;  DB 22;  Length 500;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgpargqagvmgfpgpkgaa 59


RESULT  11
AAE02703
ID   AAE02703 standard; Protein; 501 AA.
XX
AC   AAE02703;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Human alpha1 (I) type I collagen helical domain (residues 179-679).
XX
KW   Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW   encapsulant; film-forming agent; moisturising agent; thickening agent;
KW   gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW   plasma expander; colloidal volume replacement material; graft coating;
KW   medical sponge; medical plug; micro-carrier; edible composition;
KW   protein supplement; fat substitute; nutritional supplement; cell culture;
KW   edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW   cartilage degeneration; joint flexibility; food industry; beverage;
KW   alpha1 (I) type I collagen.
XX
OS   Homo sapiens.
XX
PN   WO200134646-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30791.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-329072/34.
XX
PT   Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT   prepared recombinantly -
XX
PS   Claim 21; Page 121-123; 137pp; English.
XX
CC   The patent discloses recombinant human gelatin which is useful
CC   in various compositions including binding agents, encapsulants,
CC   stabilising agents, film-forming agents, moisturising agents,
CC   emulsifiers, thickening agents, gelling agents, colloidal agents,
CC   adhesive agents, pharmaceutical compositions, hard gel capsules,
CC   soft gel capsules, plasma expander, colloidal volume replacement
CC   materials, graft coatings, medical sponges, medical plugs,
CC   pharmaceutical stabilisers, micro-carriers, edible compositions,
CC   protein supplements, fat substitutes, nutritional supplements,
CC   edible coatings, photographic compositions, cosmetic compositions,
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is human alpha1 (I) type I collagen
CC   helical domain (residues 179-679). This sequence is a recombinant
CC   gelatin.
XX
SQ   Sequence   501 AA;


  Query Match             100.0%;  Score 333;  DB 22;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 411


RESULT  12
AAB68057
ID   AAB68057 standard; Protein; 501 AA.
XX
AC   AAB68057;
```

```
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 85
FT                   /note= "this residue is given as unknown as it is
FT                          illegible in the specification"
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 114-116; 130pp; English.
XX
CC   The present sequence represents a human recombinant gelatin polypeptide.
CC   The recombinant gelatin polypeptide is used to produce vaccine
CC   formulations of the invention. The recombinant human gelatin is
CC   non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC   stability at ambient temperatures. The vaccine formulation comprises a
CC   vaccine formulated for the prevention of a disease selected from vaccinia
CC   virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC   diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC   (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC   haemophillus influenzae meningitis; rabies, cholera, Japanese
CC   encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC   adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC   respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC   herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ   Sequence   501 AA;
```

```
  Query Match             100.0%;  Score 333;  DB 22;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     353 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 411


RESULT  13
AAE02718
ID   AAE02718 standard; Protein; 662 AA.
XX
AC   AAE02718;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Human alpha1 (I) type I collagen helical domain (residues 531-1192).
XX
KW   Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW   encapsulant; film-forming agent; moisturising agent; thickening agent;
KW   gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW   plasma expander; colloidal volume replacement material; graft coating;
KW   medical sponge; medical plug; micro-carrier; edible composition;
KW   protein supplement; fat substitute; nutritional supplement; cell culture;
KW   edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW   cartilage degeneration; joint flexibility; food industry; beverage;
KW   alpha1 (I) type I collagen.
XX
OS   Homo sapiens.
XX
PN   WO200134646-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30791.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-329072/34.
XX
PT   Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT   prepared recombinantly -
XX
PS   Claim 21; Page 135-137; 137pp; English.
XX
CC   The patent discloses recombinant human gelatin which is useful
CC   in various compositions including binding agents, encapsulants,
CC   stabilising agents, film-forming agents, moisturising agents,
CC   emulsifiers, thickening agents, gelling agents, colloidal agents,
CC   adhesive agents, pharmaceutical compositions, hard gel capsules,
CC   soft gel capsules, plasma expander, colloidal volume replacement
CC   materials, graft coatings, medical sponges, medical plugs,
CC   pharmaceutical stabilisers, micro-carriers, edible compositions,
CC   protein supplements, fat substitutes, nutritional supplements,
CC   edible coatings, photographic compositions, cosmetic compositions,
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is human alpha1 (I) type I collagen
CC   helical domain (residues 531-1192). This sequence is a recombinant
CC   gelatin.
XX
SQ   Sequence   662 AA;
```

```
  Query Match             100.0%;  Score 333;  DB 22;  Length 662;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgargqagvmgfpgpkgaa 59


RESULT  14
AAB68072
ID   AAB68072 standard; Protein; 662 AA.
XX
AC   AAB68072;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
```

```
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 53
FT                   /note= "this residue is given as unknown as it is
FT                          illegible in the specification"
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 128-130; 130pp; English.
XX
CC   The present sequence represents a human recombinant gelatin polypeptide.
CC   The recombinant gelatin polypeptide is used to produce vaccine
CC   formulations of the invention. The recombinant human gelatin is
CC   non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC   stability at ambient temperatures. The vaccine formulation comprises a
CC   vaccine formulated for the prevention of a disease selected from vaccinia
CC   virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC   diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC   (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC   haemophillus influenzae meningitis, rabies, cholera, Japanese
CC   encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC   adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC   respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC   herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ   Sequence   662 AA;

  Query Match             100.0%;  Score 333;  DB 22;  Length 662;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgpargqagvmgfpgpkgaa 59


RESULT  15
AAY84541
ID   AAY84541 standard; Protein; 1057 AA.
XX
AC   AAY84541;
XX
DT   25-JUL-2000  (first entry)
XX
DE   Amino acid sequence of a human collagen 1 (alpha1) protein.
XX
KW   Extracellular matrix protein; self aggregation; hydroxylated proline;
KW   trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW   collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS   Homo sapiens.
XX
PN   EP992586-A2.
XX
PD   12-APR-2000.
XX
PF   07-OCT-1999;   99EP-0119184.
XX
PR   09-OCT-1998;   98US-0169768.
XX
PA   (USSU ) US SURGICAL CORP.
XX
PI   Gruskin EA,  Buechter DD,  Zhang G,  Connolly K;
XX
DR   WPI; 2000-259138/23.
DR   N-PSDB; AAA12502.
XX
PT   Production of extracellular matrix proteins containing
PT   4-trans-hydroxyproline results in native self aggregating proteins,
PT   useful on medical implants -
XX
PS   Disclosure; Fig 27A-E; 260pp; English.
XX
CC   The specification describes a method for producing an extracellular
CC   matrix protein or its fragment. The extracellular matrix protein is
CC   capable of self aggregating in a cell which does not ordinarily
CC   hydroxylated prolines. The method comprises optimising a nucleic acid
CC   sequence for expression in the cell by substitution of codons preferred
CC   by that cell for naturally occurring codons not preferred by the cell;
CC   incorporating the nucleic acid sequence into the cell; and contacting
CC   the cell with a hypertonic growth medium containing at least one amino
CC   acid, selected from the group consisting of trans-4-hydroxyproline and
CC   3-hydroxyproline to allow at least one of the amino acids to be
CC   assimilated into the cell and incorporated into the extracellular matrix
CC   protein. The method may be used to make host cells assimilate and
CC   incorporate trans-4-hydroxyproline into proteins. This is especially
CC   useful in the recombinant production of proteins such as collagen,
CC   fibrinogen and fibronectin whose ability to self aggregate and produce
CC   functional proteins depends on the post translational hydroxylation of
CC   proline. The method is also useful in studying the structure and function
CC   of polypeptides which do not normally contain trans-4-hydroxyproline.
CC   The present sequence represents a human collagen 1 (alpha1) protein,
CC   which may be produced using the method of the invention.
XX
SQ   Sequence   1057 AA;

  Query Match             100.0%;  Score 333;  DB 21;  Length 1057;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 eaglpgakgltgspgspgpdgktgppgpagqdgrpgppgppgpargqagvmgfpgpkgaa 428


Search completed: January 28, 2002, 07:49:43
Job time: 168 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:46:55 ; Search time 19.73 Seconds
(without alignments)
67.293 Million cell updates/sec

Title: US-09-710-239-18
Perfect score: 333
Sequence: 1 EAGLPGAKGLTGSPGSPGPD.....PPGARGQAGVMGFPGPKGAA 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	333	100.0	1057	3	US-08-931-820-1	Sequence 1, Appli
2	333	100.0	1341	3	US-08-963-825-18	Sequence 18, Appl
3	321	96.4	595	4	US-09-219-849-48	Sequence 48, Appl
4	321	96.4	595	4	US-09-219-849-50	Sequence 50, Appl
5	321	96.4	822	4	US-09-219-849-49	Sequence 49, Appl
6	256	76.9	1060	3	US-08-931-820-3	Sequence 3, Appli
7	256	76.9	1418	3	US-08-963-825-20	Sequence 20, Appl
8	256	76.9	1418	4	US-09-010-999-1	Sequence 1, Appli
9	256	76.9	1442	2	US-08-316-650-12	Sequence 12, Appl
10	256	76.9	1442	5	PCT-US95-02251-12	Sequence 12, Appl
11	234	70.3	1057	3	US-08-931-820-4	Sequence 4, Appli
12	234	70.3	1078	3	US-08-963-825-21	Sequence 21, Appl
13	209	62.8	279	4	US-09-010-999-2	Sequence 2, Appli
14	200	60.1	1024	3	US-08-931-820-2	Sequence 2, Appli
15	200	60.1	1366	3	US-08-963-825-19	Sequence 19, Appl
16	192	57.7	144	1	US-08-642-255-49	Sequence 49, Appl
17	192	57.7	234	1	US-08-642-255-51	Sequence 51, Appl
18	192	57.7	504	4	US-09-219-849-3	Sequence 3, Appli
19	192	57.7	561	1	US-08-642-255-52	Sequence 52, Appl
20	192	57.7	720	4	US-09-219-849-4	Sequence 4, Appli
21	192	57.7	777	1	US-08-642-255-53	Sequence 53, Appl
22	189	56.8	330	1	US-08-642-255-32	Sequence 32, Appl
23	189	56.8	408	1	US-07-609-716-65	Sequence 65, Appl
24	189	56.8	408	4	US-08-475-411A-65	Sequence 65, Appl
25	189	56.8	408	4	US-08-478-029A-65	Sequence 65, Appl
26	183	55.0	546	1	US-08-494-168-10	Sequence 10, Appl
27	179.5	53.9	357	1	US-07-609-716-66	Sequence 66, Appl
28	179.5	53.9	357	1	US-08-642-255-33	Sequence 33, Appl
29	179.5	53.9	357	4	US-08-475-411A-66	Sequence 66, Appl
30	179.5	53.9	357	4	US-08-478-029A-66	Sequence 66, Appl
31	179	53.8	60	1	US-08-534-342-12	Sequence 12, Appl
32	179	53.8	60	1	US-08-675-140-12	Sequence 12, Appl
33	176.5	53.0	532	1	US-08-494-168-9	Sequence 9, Appli
34	174.5	52.4	623	4	US-09-029-348-3	Sequence 3, Appli
35	174.5	52.4	626	4	US-09-029-348-2	Sequence 2, Appli
36	170.5	51.2	252	1	US-08-642-255-61	Sequence 61, Appl
37	170.5	51.2	1064	1	US-08-642-255-62	Sequence 62, Appl
38	168.5	50.6	310	4	US-09-219-849-47	Sequence 47, Appl
39	168.5	50.6	471	2	US-08-399-889-24	Sequence 24, Appl
40	168.5	50.6	471	3	US-09-167-364-24	Sequence 24, Appl
41	168.5	50.6	471	4	US-09-439-897-2	Sequence 2, Appli
42	167	50.2	446	2	US-08-836-854-15	Sequence 15, Appl
43	166.5	50.0	228	4	US-09-219-849-38	Sequence 38, Appl
44	166.5	50.0	557	3	US-09-320-095-10	Sequence 10, Appl
45	166.5	50.0	557	4	US-09-523-487-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
;   APPLICANT:
;   TITLE OF INVENTION: Assay for collagen degradation
;   NUMBER OF SEQUENCES: 4
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/931,820
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: EP 96202596.1
;     FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1057 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;     ORGANISM: Homo sapiens
;     TISSUE TYPE: Collagen type I
US-08-931-820-1


  Query Match            100.0%;  Score 333;  DB 3;  Length 1057;
  Best Local Similarity  100.0%;  Pred. No. 7.3e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 428


RESULT 2
US-08-963-825-18
; Sequence 18, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
;   APPLICANT: Qvist, Per
```

```
;    APPLICANT:  Bonde, Martin
;    TITLE OF INVENTION:  A Method for Assaying Collagen Fragments
;    TITLE OF INVENTION:  in Body Fluids, A Test Kit and Means for Carrying Out the
;    TITLE OF INVENTION:  Method and Use of the Method to Diagnose the Presence of
;    TITLE OF INVENTION:  Disorders Associated with the Metabolism of
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/963,825
;      FILING DATE:
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,319
;      FILING DATE:  21-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda C
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  4305/08701
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1341 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  COLLAGEN ALPHA 1 (I)
US-08-963-825-18


  Query Match             100.0%;  Score 333;  DB 3;  Length 1341;
  Best Local Similarity   100.0%;  Pred. No. 9.2e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     407 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 465


RESULT   3
US-09-219-849-48
; Sequence 48, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
;  APPLICANT:  VAN HEERDE, GEORGE V.
;  APPLICANT:  VAN RIJN, ALEXIS C.
;  APPLICANT:  BOUWSTRA, JAN B.
;  APPLICANT:  DE WOLF, FREDERIK A.
;  APPLICANT:  MOOBROEK, ANDREAS
;  APPLICANT:  WERTEN, MARC W.T.
;  APPLICANT:  WIND, RICHELE D.
;  APPLICANT:  VAN DEN BOSCH, TANJA J.
;  TITLE OF INVENTION:  SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
;  TITLE OF INVENTION:  SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
;  TITLE OF INVENTION:  PREPARATION THEREOF
;  FILE REFERENCE:  2728-2
;  CURRENT APPLICATION NUMBER: US/09/219,849
;  CURRENT FILING DATE:  1998-12-23
;  NUMBER OF SEQ ID NOS:  50
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 48
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: amino acid sequence
US-09-219-849-48


  Query Match             96.4%;  Score 321;  DB 4;  Length 595;
  Best Local Similarity   96.6%;  Pred. No. 5e-22;
  Matches   57;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     352 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGTA 410


RESULT   4
US-09-219-849-50
; Sequence 50, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
;  APPLICANT:  VAN HEERDE, GEORGE V.
;  APPLICANT:  VAN RIJN, ALEXIS C.
;  APPLICANT:  BOUWSTRA, JAN B.
;  APPLICANT:  DE WOLF, FREDERIK A.
;  APPLICANT:  MOOBROEK, ANDREAS
;  APPLICANT:  WERTEN, MARC W.T.
;  APPLICANT:  WIND, RICHELE D.
;  APPLICANT:  VAN DEN BOSCH, TANJA J.
;  TITLE OF INVENTION:  SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
;  TITLE OF INVENTION:  SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
;  TITLE OF INVENTION:  PREPARATION THEREOF
;  FILE REFERENCE:  2728-2
;  CURRENT APPLICATION NUMBER: US/09/219,849
;  CURRENT FILING DATE:  1998-12-23
;  NUMBER OF SEQ ID NOS:  50
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 50
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: amino acid sequence
US-09-219-849-50


  Query Match             96.4%;  Score 321;  DB 4;  Length 595;
  Best Local Similarity   96.6%;  Pred. No. 5e-22;
  Matches   57;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     352 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGTA 410


RESULT   5
US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
;  APPLICANT:  VAN HEERDE, GEORGE V.
;  APPLICANT:  VAN RIJN, ALEXIS C.
;  APPLICANT:  BOUWSTRA, JAN B.
;  APPLICANT:  DE WOLF, FREDERIK A.
;  APPLICANT:  MOOBROEK, ANDREAS
;  APPLICANT:  WERTEN, MARC W.T.
```

```
;  APPLICANT:  WIND, RICHELE D.
;  APPLICANT:  VAN DEN BOSCH, TANJA J.
;  TITLE OF INVENTION:  SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
;  TITLE OF INVENTION:  SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
;  TITLE OF INVENTION:  PREPARATION THEREOF
;  FILE REFERENCE:  2728-2
;  CURRENT APPLICATION NUMBER: US/09/219,849
;  CURRENT FILING DATE:  1998-12-23
;  NUMBER OF SEQ ID NOS:  50
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 49
;   LENGTH: 822
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: amino acid sequence
US-09-219-849-49


  Query Match             96.4%;  Score 321;  DB 4;  Length 822;
  Best Local Similarity   96.6%;  Pred. No. 6.9e-22;
  Matches   57;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||| |
Db    352 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGTA 410


RESULT   6
US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Assay for collagen degradation
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 96202596.1
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Collagen type II
US-08-931-820-3


  Query Match             76.9%;  Score 256;  DB 3;  Length 1060;
  Best Local Similarity   77.6%;  Pred. No. 6.3e-16;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || |  |:||||||||| ||||| ||||||||||
Db    372 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 429


RESULT   7
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
;  GENERAL INFORMATION:
;    APPLICANT:  Qvist, Per
;    APPLICANT:  Bonde, Martin
;    TITLE OF INVENTION:  A Method for Assaying Collagen Fragments
;    TITLE OF INVENTION:  in Body Fluids, A Test Kit and Means for Carrying Out the
;    TITLE OF INVENTION:  Method and Use of the Method to Diagnose the Presence of
;    TITLE OF INVENTION:  Disorders Associated with the Metabolism of
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/963,825
;      FILING DATE:
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,319
;      FILING DATE:  21-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda C
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  4305/08701
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1418 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  COLLAGEN -ALPHA 1 (II)
US-08-963-825-20


  Query Match             76.9%;  Score 256;  DB 3;  Length 1418;
  Best Local Similarity   77.6%;  Pred. No. 8.3e-16;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || |  |:||||||||| ||||| ||||||||||
Db    484 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 541


RESULT   8
US-09-010-999-1
; Sequence 1, Application US/09010999
; Patent No. 6132976
;  GENERAL INFORMATION:
;    APPLICANT:  Poole, Anthony R.
;    APPLICANT:  Hollander, Anthony P.
;    APPLICANT:  Billinghurst, R. C.
;    TITLE OF INVENTION:  IMMUNOASSAYS FOR THE MEASUREMENT OF
;    TITLE OF INVENTION:  COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
```

```
;       STREET:  3000 K Street, N.W., Suite 500
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/010,999
;       FILING DATE:  22-JAN-1998
;       CLASSIFICATION:  4335
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/448,501
;       FILING DATE:  17-JUL-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/984,123
;       FILING DATE:  04-DEC-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bent, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  032931/0212
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1418 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       ORGANISM:  Human Type II Collagen
US-09-010-999-1


  Query Match             76.9%;  Score 256;  DB 4;  Length 1418;
  Best Local Similarity   77.6%;  Pred. No. 8.3e-16;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | |||||:|||| ||  || || || |  |:||||||||| ||||| ||||||||||
Db     484 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 541


RESULT   9
US-08-316-650-12
; Sequence 12, Application US/08316650
; Patent No. 5942496
;  GENERAL INFORMATION:
;     APPLICANT:  Bonadio, Jeffrey
;     APPLICANT:  Roessler, Blake J.
;     APPLICANT:  Goldstein, Steven A.
;     APPLICANT:  Lin, Wushan
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS
;     TITLE OF INVENTION:  FOR STIMULATING BONE CELLS
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/316,650
;       FILING DATE:  30-SEP-1994
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/199,780
;       FILING DATE:  30-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UMIC:008
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (713) 789-2679
;       TELEX:  79-0924
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1442 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-316-650-12


  Query Match             76.9%;  Score 256;  DB 2;  Length 1442;
  Best Local Similarity   77.6%;  Pred. No. 8.5e-16;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | |||||:|||| ||  || || || |  |:||||||||| ||||| ||||||||||
Db     508 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 565


RESULT  10
PCT-US95-02251-12
; Sequence 12, Application PC/TUS9502251
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR STIMULATING BONE
;     TITLE OF INVENTION:  CELLS
;     NUMBER OF SEQUENCES:  18
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  United States of America
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;       SOFTWARE:  PatentIn Release #1.0, Version
;       SOFTWARE:  #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/02251
;       FILING DATE:  CONCURRENTLY HEREWITH
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/316,650
;       FILING DATE:  30-SEP-1994
;       CLASSIFICATION:
;       APPLICATION NUMBER:  US 08/199,780
;       FILING DATE:  18-FEB-1994
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UMIC009P--
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (713) 789-2679
```

```
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1442 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US95-02251-12


  Query Match             76.9%;  Score 256;  DB 5;  Length 1442;
  Best Local Similarity   77.6%;  Pred. No. 8.5e-16;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | ||||||:|||| ||  || || || |  |:|||||||| ||||| ||||||||||
Db    508 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 565


RESULT  11
US-08-931-820-4
; Sequence 4, Application US/08931820
; Patent No. 6010863
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Assay for collagen degradation
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 96202596.1
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1057 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Collagen type III
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1055
;      OTHER INFORMATION:  /label= Modified
;      OTHER INFORMATION:  /note= "Ala may be Pro"
US-08-931-820-4


  Query Match             70.3%;  Score 234;  DB 3;  Length 1057;
  Best Local Similarity   72.7%;  Pred. No. 6e-14;
  Matches   40;  Conservative    3;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
          | || :|: |||| || ||| |||| |: |||||||| | ||| ||||||||||
Db    383 GGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSGPRGQPGVMGFPGPKG 437


RESULT  12
US-08-963-825-21
; Sequence 21, Application US/08963825
; Patent No. 6110689
;  GENERAL INFORMATION:
;    APPLICANT:  Qvist, Per
;    APPLICANT:  Bonde, Martin
;    TITLE OF INVENTION:  A Method for Assaying Collagen Fragments
;    TITLE OF INVENTION:  in Body Fluids, A Test Kit and Means for Carrying Out the
;    TITLE OF INVENTION:  Method and Use of the Method to Diagnose the Presence of
;    TITLE OF INVENTION:  Disorders Associated with the Metabolism of
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/963,825
;      FILING DATE:
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,319
;      FILING DATE:  21-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda C
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  4305/08701
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1078 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  COLLAGEN ALPHA 1 (III)
US-08-963-825-21


  Query Match             70.3%;  Score 234;  DB 3;  Length 1078;
  Best Local Similarity   72.7%;  Pred. No. 6.1e-14;
  Matches   40;  Conservative    3;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
          | || :|: |||| || ||| |||| |: |||||||| | ||| ||||||||||
Db    384 GGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSGPRGQPGVMGFPGPKG 438


RESULT  13
US-09-010-999-2
; Sequence 2, Application US/09010999
; Patent No. 6132976
;  GENERAL INFORMATION:
;    APPLICANT:  Poole, Anthony R.
;    APPLICANT:  Hollander, Anthony P.
;    APPLICANT:  Billinghurst, R. C.
;    TITLE OF INVENTION:  IMMUNOASSAYS FOR THE MEASUREMENT OF
;    TITLE OF INVENTION:  COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
```

```
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/010,999
;       FILING DATE:  22-JAN-1998
;       CLASSIFICATION:  4335
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/448,501
;       FILING DATE:  17-JUL-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/984,123
;       FILING DATE:  04-DEC-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bent, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  032931/0212
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  279 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     ORIGINAL SOURCE:
;       ORGANISM:  CB11 peptide of Tpye II Collagen
US-09-010-999-2


  Query Match             62.8%;  Score 209;  DB 4;  Length 279;
  Best Local Similarity   74.0%;  Pred. No. 3e-12;
  Matches   37;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVM 50
          | |||||:|||| ||  || || || || |  |:|||||||||| ||||| |||
Db    230 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVM 279


RESULT  14
US-08-931-820-2
; Sequence 2, Application US/08931820
; Patent No. 6010863
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Assay for collagen degradation
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 96202596.1
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1024 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Collagen type I
US-08-931-820-2


  Query Match             60.1%;  Score 200;  DB 3;  Length 1024;
  Best Local Similarity   66.7%;  Pred. No. 6.7e-11;
  Matches   38;  Conservative    4;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
          | || | :|| |||: || || || |  | |||||| || ||||: | :|||||||
Db    364 EPGLMGPRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPVGARGEPGNIGFPGPKG 420


RESULT  15
US-08-963-825-19
; Sequence 19, Application US/08963825
; Patent No. 6110689
;  GENERAL INFORMATION:
;    APPLICANT:  Qvist, Per
;    APPLICANT:  Bonde, Martin
;    TITLE OF INVENTION:  A Method for Assaying Collagen Fragments
;    TITLE OF INVENTION:  in Body Fluids, A Test Kit and Means for Carrying Out the
;    TITLE OF INVENTION:  Method and Use of the Method to Diagnose the Presence of
;    TITLE OF INVENTION:  Disorders Associated with the Metabolism of
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/963,825
;      FILING DATE:
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,319
;      FILING DATE:  21-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda C
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  4305/08701
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1366 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  collagen alpha 2- type I
US-08-963-825-19


  Query Match             60.1%;  Score 200;  DB 3;  Length 1366;
  Best Local Similarity   66.7%;  Pred. No. 8.9e-11;
  Matches   38;  Conservative    4;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
          | || | :|| |||: || || || |  | |||||| || ||||: | :|||||||
```

Db 443 EPGLMGPRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPVGARGEPGNIGFPGPKG 499

Search completed: January 28, 2002, 07:48:59
Job time: 124 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:46:55 ; Search time 21.88 Seconds
(without alignments)
205.407 Million cell updates/sec

Title: US-09-710-239-18
Perfect score: 333
Sequence: 1 EAGLPGAKGLTGSPGSPGPD.....PPGARGQAGVMGFPGPKGAA 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	333	100.0	1464	1	CGHU1S	collagen alpha 1(I
2	325	97.6	1042	1	CGCH1S	collagen alpha 1(I
3	321	96.4	1453	2	S21626	collagen alpha 1(I
4	317	95.2	671	1	CGRT1S	collagen alpha 1(I
5	261	78.4	1492	2	A40333	collagen alpha 1'(
6	258	77.5	1486	1	B40333	collagen alpha 1(I
7	257	77.2	673	1	CGBO6C	collagen alpha 1(I
8	256	76.9	1418	2	T45467	collagen alpha 1(I
9	256	76.9	1419	2	A41182	collagen alpha 1(I
10	256	76.9	1487	1	CGHU6C	collagen alpha 1(I
11	256	76.9	1487	2	B41182	collagen alpha 1(I
12	234	70.3	1466	1	CGHU7L	collagen alpha 1(I
13	233	70.0	1464	2	S59856	collagen alpha 1(I
14	232	69.7	886	2	I50694	collagen alpha 1(I
15	228	68.5	1049	1	CGBO7S	collagen alpha 1(I
16	225	67.6	1497	2	I49607	procollagen type V
17	224	67.3	1496	1	CGHU2V	collagen alpha 2(V
18	204	61.3	1373	1	A43291	collagen alpha 2(I
19	200	60.1	1366	1	CGHU2S	collagen alpha 2(I
20	191	57.4	1763	2	S16366	collagen alpha 2(I
21	186.5	56.0	184	1	CGRT2S	collagen alpha 2(I
22	183	55.0	296	2	A31219	collagen 1 - Caeno
23	183	55.0	301	2	T21314	hypothetical prote
24	183	55.0	1691	1	S22917	collagen alpha 5(I
25	181.5	54.5	1669	1	CGMS4B	collagen alpha 1(I
26	181	54.4	488	2	A27353	collagen alpha 1(I
27	180.5	54.2	178	2	A39762	collagen alpha 1(X
28	180	54.1	675	2	S20819	collagen alpha 3(I
29	180	54.1	1690	1	CGHU1B	collagen alpha 4(I
30	179.5	53.9	674	2	S13301	collagen alpha 1(X
31	179.5	53.9	680	1	CGHU1D	collagen alpha 1(X
32	178	53.5	303	2	T19289	hypothetical prote
33	176.5	53.0	302	2	T15936	hypothetical prote
34	176.5	53.0	1669	1	CGHU4B	collagen alpha 1(I
35	176	52.9	325	2	S02170	collagen alpha 1(I
36	176	52.9	1414	1	S23809	collagen alpha 2(I
37	175.5	52.7	754	2	A55267	collagen alpha 5(I
38	175	52.6	296	2	T24827	hypothetical prote
39	174.5	52.4	632	2	S42731	collagen alpha 1 c
40	174	52.3	295	2	A44984	collagen - nematod
41	173.5	52.1	636	2	S41067	collagen alpha 1(I
42	173.5	52.1	1315	2	A56101	collagen alpha 1(X
43	173.5	52.1	1774	2	B56101	collagen alpha 1(X
44	173.5	52.1	1838	1	CGHU1V	collagen alpha 1(V
45	173.5	52.1	1843	2	S18803	collagen alpha 1(V

ALIGNMENTS

RESULT 1
CGHU1S
collagen alpha 1(I) chain precursor - human
N;Alternate names: procollagen alpha 1(I) chain
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 31-Dec-2000
C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; 5269; A29439; I53466; A02852; I37247
R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorious, P
Gene 67, 105-115, 1988
A;Title: Complete nucleotide sequence of the region encompassing the first twenty-fiv
A;Reference number: I60114; MUID:88329734
A;Accession: I60114
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-369,'L',371-589 <DAL>
A;Cross-references: GB:M20789; NID:g179593; PIDN:AAB59373.1; PID:g179594
R;Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Pr
Biochem. J. 253, 919-922, 1988
A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of hum
A;Reference number: S01143; MUID:89025644
A;Accession: S01143
A;Molecule type: mRNA
A;Residues: 1-472 <TRO>
A;Cross-references: EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID:g30016; GB:M36546;
A;Note: submitted to the EMBL/GenBank/DDBJ databases by Prockop, D.J., 13-JUN-1988
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams,
Nature 310, 337-340, 1984
A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation
A;Reference number: A93335; MUID:84270697
A;Accession: A93335
A;Molecule type: DNA
A;Residues: 1-58,'Q',60-181 <CHU>
A;Cross-references: EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID:g35658
R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet,
J. Biol. Chem. 262, 15151-15157, 1987
A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene
A;Reference number: I55254; MUID:88033098
A;Accession: I55254
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-45 <ROS>
A;Cross-references: GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:g180388
R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A;Title: Regulatory elements in the first intron contribute to transcriptional contro
A;Reference number: A39943; MUID:88097389
A;Accession: A39943
A;Molecule type: DNA
A;Residues: 1-34 <BOR>
A;Cross-references: GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:g553238
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A;Reference number: I55237; MUID:85130970
A;Accession: I55237
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-34 <CH2>
A;Cross-references: GB:M10627; NID:g180383; PIDN:AAA51992.1; PID:g553226
R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A;Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
rome, type VII.
A;Reference number: A35233; MUID:90202908
A;Accession: A35233
A;Molecule type: protein
A;Residues: 33-52 <WIR>
A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A;Title: A base substitution in the exon of a collagen gene causes alternative splicing
A;Reference number: S09400; MUID:89356643
A;Accession: S09400
A;Molecule type: mRNA
A;Residues: 156-183 <WEI>
R;Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alphal
A;Reference number: A90567; MUID:71038625
A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A;Accession: B90567
A;Molecule type: protein
A;Residues: 162-198,'Z',200-201,'Z',203-206,'Z',208-209,'Z',211-228,'B',230,'BB',233,'Z'
A;Experimental source: skin
A;Note: evidence for 170-allysine
R;Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, P
Eur. J. Biochem. 192, 153-159, 1990
A;Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A;Reference number: S11372; MUID:90382436
A;Accession: S11372
A;Molecule type: protein
A;Residues: 175-187;274-287,'P',289 <BAE>
A;Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A;Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
ooperative melting of intact type I collagen.
A;Reference number: I55342; MUID:92042092
A;Accession: I55342
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 258-268;1347-1357 <DEA>
A;Cross-references: GB:S67495; NID:g239007; PIDN:AAB20350.1; PID:g239008
A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A;Reference number: A92069; MUID:71001508
A;Accession: A92069
A;Molecule type: protein
A;Residues: 263-268 <MOR>
A;Experimental source: skin
A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R;Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A;Title: Segmental amplification of the entire helical and telopeptide regions of the cD
A;Reference number: S15989; MUID:90326017
A;Accession: S15989
A;Molecule type: mRNA
A;Residues: 281-302;402-420;823-843;925-944;1026-1045;1143-1162 <LAB>
R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
Connect. Tissue Res. 29, 1-11, 1993
A;Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A;Reference number: I52905; MUID:93339042
A;Accession: I52905
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 342-352,'C',354-359 <WI2>
A;Cross-references: GB:S64717; NID:g408195; PIDN:AAB27677.1; PID:g408196
A;Note: mutant sequence from patient with osteogenesis imperfecta
R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A;Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalph
A;Reference number: A90476; MUID:84080385
A;Accession: A90476
A;Molecule type: mRNA
A;Residues: 425-1250,'X',1252-1328,'S',1330-1390,'X',1392-1464 <BER>
A;Cross-references: GB:K01228; NID:g180391; PIDN:AAA51995.1; PID:g180392
A;Note: sequence partially completed for missing nucleotides by A29439
R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A;Title: Multiexon deletion in an osteogenesis imperfecta variant with increased type
A;Reference number: A22161; MUID:85104934
A;Accession: A22161
A;Molecule type: DNA
A;Residues: 472-594,'R',596-607 <CH3>
A;Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PIDN:AAA51847.1;
A;Note: the authors translated the codon CGT for residue 595 as Pro
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explai
.
A;Reference number: A35336; MUID:90252792
A;Accession: A35336
A;Molecule type: mRNA
A;Residues: 710-720,'E',722-737,'E',739-745 <WAL>
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mot
Hum. Mol. Genet. 3, 2201-2206, 1994
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in th
A;Reference number: I54365; MUID:95187161
A;Accession: I54365
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 746-766,'S',768-781 <FOR>
A;Cross-references: GB:L47667; NID:g1009093; PIDN:AAB59576.1; PID:g1009094
R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of
A;Reference number: A47426; MUID:93352646
A;Accession: A47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1336,1339-1387,'R',1389-1464 <CHE>
A;Cross-references: GB:S64596; NID:g407589; PIDN:AAB27856.1; PID:g407590
A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIP:136445)
A;Note: does not represent an experimentally determined sequence but three different
A;Accession: B47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH4>
A;Experimental source: normal dermal fibroblast culture
A;Accession: C47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1464 <CH5>
A;Experimental source: fetal cell 86-237
A;Accession: D47426
A;Molecule type: mRNA
A;Residues: 1179-1336,1339-1464 <CH6>
A;Experimental source: fetal cell 86-146
A;Accession: E47426
A;Molecule type: mRNA
A;Residues: 1179-1387,'R',1389-1464 <CH7>
A;Experimental source: fetal cell 88-251
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.;
J. Biol. Chem. 263, 14605-14607, 1988
A;Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptid
A;Reference number: I55269; MUID:89008319
A;Accession: I55269

```
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1187-1194,'C',1196-1220 <COH>
A;Cross-references: GB:M23213; NID:g340842; PIDN:AAB59363.1; PID:g499622
A;Note: mutant sequence from a patient with mild osteogenesis imperfecta
R;Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.
Nucleic Acids Res. 16, 349, 1988
A;Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

  Query Match            100.0%;  Score 333;  DB 1;  Length 1464;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-20;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 589


RESULT   2
CGCH1S
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
C;Accession: A90458; A90181; A02857
R;Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A;Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete prima
A;Reference number: A90458; MUID:82231995
A;Accession: A90458
A;Molecule type: protein
A;Residues: 1-1036 <HIG>
A;Experimental source: skin
A;Note: this is the latest in a series of papers from these workers elucidating the sequ
R;Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
A;Reference number: A90181; MUID:72243016
A;Accession: A90181
A;Molecule type: protein
A;Residues: 1037-1042 <EYR>
A;Experimental source: skin
A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
C;Comment: Most of the prolines at the third position of the tripeptide repeating unit (
C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pc
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental


  Query Match            97.6%;  Score 325;  DB 1;  Length 1042;
  Best Local Similarity  98.3%;  Pred. No. 5.4e-20;
  Matches   58;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    369 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGAA 427


RESULT   3
S21626
collagen alpha 1(I) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 25-Apr-1997 #text_change 13-Aug-1999
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R;Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A;Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A;Reference number: S57243
A;Accession: S57243
A;Molecule type: mRNA
A;Residues: 1-1453 <LIS>
A;Cross-references: EMBL:U08020; NID:g470673; PIDN:AAA88912.1; PID:g470674
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNA
A;Reference number: S16176; MUID:91274355
A;Accession: S16374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1453 <MET>
A;Cross-references: EMBL:X57981; NID:g50484; PIDN:CAA41046.1; PID:g50485
R;French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A;Reference number: A23982; MUID:86137403
A;Accession: A23982
A;Molecule type: mRNA
A;Residues: 518-1128 <FRE>
A;Cross-references: GB:M14423; NID:g192261; PIDN:AAA37333.1; PID:g192262
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A;Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence f
A;Reference number: I49559; MUID:83141374
A;Accession: I49559
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 735-1130 <RES>
A;Cross-references: GB:M17491; NID:g192263; PIDN:AAA37334.1; PID:g192264
R;Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A;Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene lea
A;Reference number: I49557; MUID:84170331
A;Accession: I49557
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-25 <RE2>
A;Cross-references: GB:K01688; NID:g192246; PIDN:AAA37330.1; PID:g553881
R;Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A;Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A;Reference number: S39789; MUID:94092741
A;Accession: S39789
A;Molecule type: DNA
A;Residues: 1-80,'E',82-105,'D',107-185;1031-1201,'G',1203-1218,'E',1220-1221,'T',122
R;Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A;Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indi
A;Reference number: I48300; MUID:94344105
A;Accession: I48300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-80,'E',82-105,'D',107-147 <REF>
A;Cross-references: EMBL:X54876; NID:g50486; PIDN:CAA38657.1; PID:g50487
C;Genetics:
A;Gene: COL1A1
A;Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 10
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple hel
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F;30-89/Domain: von Willebrand factor type C repeat homology <VWC>
F;152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
F;1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match            96.4%;  Score 321;  DB 2;  Length 1453;
  Best Local Similarity  96.6%;  Pred. No. 1.5e-19;
  Matches   57;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||| |||||||||||||||||| |
Db    520 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGTA 578


RESULT   4
```

CGRT1S
collagen alpha 1(I) chain - rat (tentative sequence) (fragments)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C;Accession: A90559; A90552; A92029; A90353; A90566; A90357; A90362; A90379; A91209; A91
R;Bornstein, P.
Biochemistry 8, 63-71, 1969
A;Title: Comparative sequence studies of rat skin and tendon collagen. II. The absence c
A;Reference number: A90559; MUID:69155173
A;Contents: CNBr0 and CNBr1
A;Accession: A90559
A;Molecule type: protein
A;Residues: 1-19 <BO1>
A;Experimental source: tendon
A;Note: sequences from skin and tendon appear to be identical
A;Note: the amino-terminal tetrapeptide may be removed by limited proteolysis during ext
R;Kang, A.H.; Bornstein, P.; Piez, K.A.
Biochemistry 6, 788-795, 1967
A;Title: The amino acid sequence of peptides from the cross-linking region of rat skin c
A;Reference number: A90552; MUID:67162268
A;Contents: CNBr1
A;Accession: A90552
A;Molecule type: protein
A;Residues: 5-19 <KAN>
A;Experimental source: skin
R;Bornstein, P.
J. Biol. Chem. 242, 2572-2574, 1967
A;Title: The incomplete hydroxylation of individual prolyl residues in collagen.
A;Reference number: A92029; MUID:67165368
A;Contents: CNBr2
A;Accession: A92029
A;Molecule type: protein
A;Residues: 20-55 <BO2>
A;Experimental source: skin and tendon
R;Butler, W.T.; Ponds, S.L.
Biochemistry 10, 2076-2081, 1971
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90353; MUID:71263178
A;Contents: CNBr4
A;Accession: A90353
A;Molecule type: protein
A;Residues: 56-102 <BU1>
A;Experimental source: skin
R;Butler, W.T.
Biochemistry 9, 44-50, 1970
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The cov
A;Reference number: A90566; MUID:70085124
A;Contents: CNBr5
A;Accession: A90566
A;Molecule type: protein
A;Residues: 103-139 <BU2>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxyla
A;Reference number: A90357; MUID:72136131
A;Contents: CNBr8
A;Accession: A90357
A;Molecule type: protein
A;Residues: 140-238 <BA1>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxyla
A;Reference number: A90362; MUID:73006942
A;Contents: CNBr8
A;Accession: A90362
A;Molecule type: protein
A;Residues: 239-418 <BA2>
A;Experimental source: skin
R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2946-2953, 1974
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90379; MUID:74271984
A;Contents: CNBr3
A;Accession: A90379
A;Molecule type: protein
A;Residues: 419-567 <BU3>
A;Experimental source: skin
R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in ne
A;Reference number: A91209; MUID:74011954
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <ST1>
A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chai
A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen)
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha1-chain.
A;Reference number: A91385; MUID:73049495
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <ST2>
A;Experimental source: skin
A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the seq
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1
C;Comment: Prolines and lysines at the third position of the tripeptide repeating uni
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin co
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydro
F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #st
F;9/Modified site: allysine (Lys) #status experimental
F;103,424,547/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (Lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 95.2%; Score 317; DB 1; Length 671;
Best Local Similarity 93.2%; Pred. No. 1.7e-19;
Matches 55; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           |||||||||||||||||||||||||||||::||||| |||||||||||||||||| |
Db     369 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGZBGRPGPAGPPGARGQAGVMGFPGPKGTA 427
```

RESULT 5
A40333
collagen alpha 1'(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C;Accession: A40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis
A;Reference number: A40333; MUID:92011898
A;Accession: A40333
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1492 <SUA>
A;Cross-references: GB:M63596
A;Note: this sequence is presented as substitutions relative to another sequence in a
es they replace; the appropriate interpretation of the sequence figure was reconstruc
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VWC>
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 78.4%; Score 261; DB 2; Length 1492;
Best Local Similarity 79.3%; Pred. No. 1.5e-14;
Matches 46; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

```
Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || | ||:|||||||| ||||| ||||||||||
Db    558 EPGLPGARGLTGRPGDAGPQGKVGPSGAAGEDGRPGPPGPQGARGQPGVMGFPGPKGA 615
```

RESULT 6
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: B40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A;Reference number: A40333; MUID:92011898
A;Accession: B40333
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1486 <SUA>
A;Cross-references: GB:M63595
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VWC>
F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 77.5%; Score 258; DB 1; Length 1486;
Best Local Similarity 77.6%; Pred. No. 2.6e-14;
Matches 45; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

```
Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || | :|:|||||||| ||||| ||||||||||
Db    555 EPGLPGARGLTGRPGDAGPQGKVGPSGASGEDGRPGPPGPQGARGQPGVMGFPGPKGA 612
```

RESULT 7
CGBO6C
collagen alpha 1(II) chain precursor - bovine (tentative sequence) (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 17-May-1996 #text_change 31-Mar-2000
C;Accession: A90369; A90396; A92210; S03940; A90189; A05039; A02859
R;Miller, E.J.; Lunde, L.G.
Biochemistry 12, 3153-3159, 1973
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A;Reference number: A90369; MUID:73258693
A;Contents: composition of CNBr1 and CNBr4
A;Accession: A90369
A;Molecule type: protein
A;Residues: 1-15 <MIL>
A;Experimental source: cartilage
A;Note: residues positioned by comparison with human alpha 1(II) chain
R;Butler, W.T.; Miller, E.J.; Finch Jr., J.E.
Biochemistry 15, 3000-3006, 1976
A;Title: The covalent structure of cartilage collagen. Amino acid sequence of the NH-2-t
A;Reference number: A90396; MUID:76253504
A;Contents: fragments CNBr2 (16-18), CNBr3 (19-21), CNBr6 (22-54), CNBr12 (55-138), and
A;Accession: A90396
A;Molecule type: protein
A;Residues: 16-177 <BUT>
A;Experimental source: cartilage
A;Note: order of CNBr peptides determined
R;Butler, W.T.; Finch Jr., J.E.; Miller, E.J.
J. Biol. Chem. 252, 639-643, 1977
A;Title: The covalent structure of cartilage collagen. Evidence for sequence heterogenei
A;Reference number: A92210; MUID:77093864
A;Accession: A92210
A;Molecule type: protein
A;Residues: 139-178,'Z',180-184,'PA',187-190,'AS',193-194,'T',196-198 <BU2>
A;Experimental source: cartilage
A;Note: a minor, probably nonallelic, alpha 1(II) component has 143-Ala, 164-Leu, and
R;Seyer, J.M.; Hasty, K.A.; Kang, A.H.
Eur. J. Biochem. 181, 159-173, 1989
A;Title: Covalent structure of collagen. Amino acid sequence of an arthritogenic cyan
A;Reference number: S03940; MUID:89231683
A;Accession: S03940
A;Molecule type: protein
A;Residues: 139-417 <SEY>
R;Butler, W.T.; Miller, E.J.; Finch Jr., J.E.; Inagami, T.
Biochem. Biophys. Res. Commun. 57, 190-195, 1974
A;Title: Homologous regions of collagen alpha1 (I) and alpha1(II) chains: apparent cl
A;Reference number: A90189; MUID:74163168
A;Accession: A90189
A;Molecule type: protein
A;Residues: 418-492 <BU3>
A;Experimental source: cartilage
A;Note: the first 75 residues of CNBr8, which follows CNBr11
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Ramirez, F.
Nucleic Acids Res. 13, 2815-2826, 1985
A;Reference number: A05039; MUID:85215651
A;Accession: A05039
A;Molecule type: mRNA
A;Residues: 493-673 <SAN>
A;Cross-references: GB:X02420; NID:g265; PIDN:CAA26269.1; PID:g266
A;Experimental source: cartilage
C;Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) ar
C;Comment: Type II collagen molecules are trimers of identical alpha 1(II) chains, ge
the CNBr peptides was determined as 1-4-2-3-6-12-11-8-10-5-9,7-14-15.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: cartilage; coiled coil; extracellular matrix; glycoprotein; hydroxylysine
F;493-673/Domain: carboxyl-terminal propeptide (fragment) #status predicted <CTP>
F;499-673/Domain: fibrillar collagen carboxyl-terminal homology (fragment) <FCC>
F;9,102,114,123,189,423,435/Modified site: 5-hydroxylysine (Lys) #status experimental
F;9,102,114,123,189,423,435/Binding site: carbohydrate (Lys) (covalent) #status exper
F;574/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 77.2%; Score 257; DB 1; Length 673;
Best Local Similarity 79.3%; Pred. No. 1.6e-14;
Matches 46; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

```
Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          |||||||||||| ||  || || || |  |:|||||||| ||||| || ||||||||
Db    368 EAGLPGAKGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVHGFPGPKGA 425
```

RESULT 8
T45467
collagen alpha 1(II) chain precursor [imported] - horse
N;Alternate names: type II collagen
C;Species: Equus caballus (domestic horse)
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
C;Accession: T45467
R;Richardson, D.W.; Dodge, G.R.
submitted to the EMBL Data Library, June 1996
A;Description: Cloning of equine type II collagen and modulation of its expression in
A;Reference number: Z22977
A;Accession: T45467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1418 <RIC>
A;Cross-references: EMBL:U62528; PIDN:AAB05773.1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 76.9%; Score 256; DB 2; Length 1418;
Best Local Similarity 77.6%; Pred. No. 3.7e-14;
Matches 45; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

```
Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || |  |:|||||||| ||||| ||||||||||
Db    484 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 541
```

```
RESULT   9
A41182
collagen alpha 1(II) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C;Accession: A41182; A44885
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A;Reference number: A41182; MUID:91358489
A;Accession: A41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1419 <MET>
A;Cross-references: GB:M65161
R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A;Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilag
A;Reference number: A44885; MUID:91347939
A;Accession: A44885
A;Molecule type: DNA
A;Residues: 1-28 <CHE>
A;Cross-references: GB:S63190; NID:g234368; PIDN:AAB19627.1; PID:g234369
A;Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBIP:63192)
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F;1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             76.9%;  Score 256;  DB 2;  Length 1419;
  Best Local Similarity   77.6%;  Pred. No. 3.7e-14;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | ||||:|||| ||  || || || |  |:|||||||| ||||| ||||||||||
Db     485 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 542


RESULT  10
CGHU6C
collagen alpha 1(II) chain precursor [validated] - human
N;Alternate names: procollagen alpha 1(II) chain
N;Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 08-Dec-2000
C;Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63
7250; I37251; I37252; I37253; I37254; I55338; I59535; I61910
R;Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A;Title: The human type II procollagen gene: identification of an additional protein-cod
A;Reference number: A38513; MUID:91184811
A;Accession: A38513
A;Molecule type: DNA
A;Residues: 1-103 <RYA>
A;Cross-references: GB:M60299; NID:g180883; PIDN:AAA73873.1; PID:g180884
R;Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A;Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A;Reference number: S06715; MUID:90067946
A;Accession: S06715
A;Molecule type: mRNA
A;Residues: 1-28,'R',99-1487 <SU2>
A;Cross-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID:g29516
A;Note: alternative splice form 1
R;Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
Biochem. J. 285, 287-294, 1992
A;Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A;Reference number: S24270; MUID:92344585
A;Accession: S24270
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-28 <VIK>
A;Cross-references: EMBL:X58709; GB:S40537; NID:g35659
A;Note: this translation is not annotated in GenBank entry HSPROCOE1, release 111.0
R;Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
Gene 44, 11-16, 1986
A;Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A;Reference number: A24828; MUID:87031574
A;Accession: A24828
A;Molecule type: DNA
A;Residues: 1-8,'T',10-28 <NUN>
A;Cross-references: GB:M25698; NID:g180872; PIDN:AAA52051.1; PID:g553237
R;Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A;Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(I
A;Reference number: S06496; MUID:90026318
A;Accession: S06496
A;Molecule type: mRNA
A;Residues: 7-28,'R',99-157,'P',159-440,'G',442-456,'E',458-640,'A',642-831,'PA',834,
A;Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
A;Note: alternative splice form 1
R;Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A;Title: Differential expression of a cysteine-rich domain in the amino-terminal prop
A;Reference number: A35428; MUID:90285153
A;Accession: A35428
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-81,'L',83-103 <RYA2>
A;Note: alternative splice form 2; splicing appears to be under developmental regulat
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A;Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide c
A;Reference number: A30147; MUID:89233138
A;Accession: A30147
A;Molecule type: DNA
A;Residues: 104-157,'P',159-236 <SUM>
A;Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168;
R;Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A;Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of
A;Reference number: A94227; MUID:90370826
A;Accession: A33116
A;Molecule type: DNA
A;Residues: 171-172,'C',174-175 <ALA>
A;Note: mutant sequence from a family with family with primary generalized osteoarthr
R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecula
A;Reference number: S64673; MUID:96195147
A;Accession: S64674
A;Molecule type: protein
A;Residues: 188-189,'X',191-195;1224-1230,'X',1232-1236 <DIA>
R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbag
Eur. J. Biochem. 234, 125-131, 1995
A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil
A;Reference number: S63514; MUID:96096730
A;Accession: S63514
A;Molecule type: protein
A;Residues: 243-261;575-590;756-763,'X',765-779 <FRA>
R;Tiller, G.E.; Weis, M.A.; Polumbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Ey
Am. J. Hum. Genet. 56, 388-395, 1995
A;Title: An RNA-splicing mutation (G+5IVS20) in the type II collagen gene (COL2A1) in
A;Reference number: I38867; MUID:95150028
A;Accession: I38867
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TIL1>
A;Cross-references: EMBL:U15195; NID:g557053; PIDN:AAB60370.1; PID:g557054
R;Ramirez, F.
submitted to the EMBL Data Library, December 1988
A;Reference number: S04892
A;Accession: S04892
A;Molecule type: mRNA
A;Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'F',836-1214 <RAM>
```

A;Cross-references: EMBL:X13783; NID:g30037; PIDN:CAA32030.1; PID:g930050
R;Vikkula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A;Title: Structural analyses of the polymorphic area in type II collagen gene.
A;Reference number: S05000; MUID:89325561
A;Accession: S05000
A;Molecule type: DNA
A;Residues: 630-640,'A',642-785 <VIK2>
A;Cross-references: EMBL:X16158; NID:g29951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
J. Biol. Chem. 267, 22522-22526, 1992
A;Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain prc
A;Reference number: A44309; MUID:93054548
A;Accession: A44309
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA; mRNA
A;Residues: 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',1038-1052,'E',1054-1068,'T',
A;Cross-references: GB:L00977; NID:g180812; PIDN:AAB23914.1; PID:g258774
A;Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence wer
A;Note: this translation is not annotated and this publication is not cited in GenBank e
A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individua
A;Reference number: S16502; MUID:90251662
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184,'GPSGKDGANGIPGPI',1185-1199 <TIL2>
A;Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>
A;Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A;Reference number: A27280; MUID:88067771
A;Accession: A27280
A;Molecule type: DNA; mRNA
A;Residues: 1175-1487 <ELI>
A;Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;1395-1408 <VAN>
A;Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prop
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human typ
A;Reference number: A21733; MUID:84118798
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295 <STR1>
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909,'PE' <STR2>
A;Cross-references: GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human prc
gene.
A;Reference number: A24561; MUID:86104139
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358 <NUN2>
A;Cross-references: GB:M12048; NID:g180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A;Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramir
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) c
A;Reference number: I37249; MUID:85215609
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28;'R',99-114;541-578;786-802;1055-1056,'N',1058-1068,'T',1070-1109;120
A;Accession: I84453
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 7-28 <SAN2>
A;Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I37250
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 541-560 <SAN3>
A;Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A;Accession: I37251

Query Match 76.9%; Score 256; DB 1; Length 1487;
Best Local Similarity 77.6%; Pred. No. 3.8e-14;
Matches 45; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
Db 553 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 610

RESULT 11
B41182
collagen alpha 1(II) chain precursor (long splice form) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Jul-1999
C;Accession: B41182
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, a
A;Reference number: A41182; MUID:91358489
A;Accession: B41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1487 <MET>
A;Cross-references: GB:M65161
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; tr
F;33-91/Domain: von Willebrand factor type C repeat homology <VWC>
F;1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 76.9%; Score 256; DB 2; Length 1487;
Best Local Similarity 77.6%; Pred. No. 3.8e-14;
Matches 45; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
Db 553 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 610

RESULT 12
CGHU7L
collagen alpha 1(III) chain precursor - human
N;Alternate names: procollagen alpha 1(III) chain
C;Species: Homo sapiens (man)
C;Date: 24-Apr-1984 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
C;Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511;
R;Prockop, D.J.

submitted to the EMBL Data Library, February 1989
A;Reference number: S05272
A;Accession: S05272
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1240,'V',1242-1466 <PRC>
A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
R;Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A;Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of huma
erences.
A;Reference number: S04642; MUID:89350838
A;Accession: S04642
A;Molecule type: mRNA
A;Residues: 1-1196 <ALA>
A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
A;Note: the complete sequence is not shown
R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (
A;Reference number: PE0011; MUID:89378752
A;Accession: PE0011
A;Molecule type: DNA
A;Residues: 1-176 <BEN>
A;Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814
R;Toman, P.D.; Ricca, G.A.; de Crombrugghe, B.
Nucleic Acids Res. 16, 7201, 1988
A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A;Reference number: S01726; MUID:88303360
A;Accession: S01726
A;Molecule type: mRNA
A;Residues: 1-170 <TOM>
A;Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061
A;Note: the authors translated the codon CAG for residue 154 as His
R;Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A;Reference number: S04887; MUID:89386015
A;Accession: S04887
A;Molecule type: mRNA
A;Residues: 149-163,'G',164-240,'D',242-471,'D',473-487,'L',489,'S',491-613,'Y',615-634,
A;Cross-references: EMBL:X15332; NID:g29545; PIDN:CAA33387.1; PID:g930045
A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R;Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A;Reference number: A90399; MUID:77134724
A;Accession: A90399
A;Molecule type: protein
A;Residues: 'V',169-225,229-232,'P',234-292,'D',294-398 <SEY1>
A;Experimental source: liver
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
R;Seyer, J.M.
submitted to the Atlas, December 1977
A;Reference number: A94562
A;Accession: A94562
A;Molecule type: protein
A;Residues: 'V',169-225,229-277,'A',279-292,'D',294,'S',296-398 <SEY2>
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A;Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
fspring.
A;Reference number: I51868; MUID:93304430
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 186-194 <MIL>
A;Cross-references: GB:S62925; NID:g386425; PIDN:AAD13937.1; PID:g4261637
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3

A;Reference number: S59511; MUID:96067614
A;Accession: S59511
A;Molecule type: mRNA
A;Residues: 302-423 <CHI>
A;Cross-references: GB:S79877; NID:g1195576; PIDN:AAB35615.1; PID:g1195577
R;Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr
A;Reference number: A90414; MUID:79000343
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675,'N',677-727 <SEY3>
A;Experimental source: liver
R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of
A;Reference number: I55349; MUID:91161621
A;Accession: I55349
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: GB:M59312; NID:g180815; PIDN:AAA52041.1; PID:g180816
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from
A;Reference number: A90438; MUID:80198282
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895,'A',897-964 <SEY4>
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; C
J. Biol. Chem. 265, 17070-17077, 1990
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping
A;Reference number: A38303; MUID:91009133
A;Accession: A38303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:g180878; PIDN:AAB59383.1; PI
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos
R;Mankoo, B.S.; Dalgleish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018,'Y',1020-1183,'S',1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID:g30054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from
A;Reference number: A90446; MUID:81208139
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979,'A',981-984,'PS',987,'QN',990-1096,'P',1098-1152,'AT',1155,'S',11
A;Experimental source: liver
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.;
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procoll
A;Reference number: A93551; MUID:85087944
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155,'P',1157-1466 <LOI>
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:g29584; PIDN:CAA25821.1
R;Miskulin, M.; Dalgleish, R.; Kluve-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Br
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the t
A;Reference number: I52393; MUID:86187804
A;Accession: I52393
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MIS>
A;Cross-references: GB:M13146; NID:g180415; PIDN:AAA52003.1; PID:g180416

```
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505
A;Accession: I79359
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: GB:M11134; NID:g180417; PIDN:AAA52004.1; PID:g180418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. P
A;Reference number: A92516; MUID:85157600
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240,'V',1242-1356,'P',1358-1466 <CHU>
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C
C;Genetics:
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
er of their length, is formed with desmosine cross-links made from lysine and allysine r
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains inte
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>
F;31-91/Domain: von Willebrand factor type C repeat homology <VWC>
F;154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;154-167/Region: amino-terminal nonhelical telopeptide
F;168-1196/Region: helical
F;1091-1093/Region: cell attachment (R-G-D) motif
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F;154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
F;161,1212/Modified site: allysine (Lys) #status predicted
F;263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
F;263/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F;948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
F;1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

  Query Match             70.3%;  Score 234;  DB 1;  Length 1466;
  Best Local Similarity   72.7%;  Pred. No. 2.5e-12;
  Matches   40;  Conservative    3;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           | || :|: |||| || ||| ||||  |: |||||||| | ||| ||||||||||
Db     531 GGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSGPRGQPGVMGFPGPKG 585


RESULT  13
S59856
collagen alpha 1(III) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 13-Aug-1999
C;Accession: S59856; S62120; S16373
R;Toman, P.D.; de Crombrugghe, B.
Gene 147, 161-168, 1994
A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
A;Reference number: S59856; MUID:95011609
A;Accession: S59856
A;Molecule type: DNA
A;Residues: 1-1464 <TOM>
A;Cross-references: EMBL:X52046
R;Toman, D.
submitted to the EMBL Data Library, November 1994
A;Reference number: S62120
A;Accession: S62120
A;Molecule type: DNA
A;Residues: 1-866,'G',868-1464 <TOA>
A;Cross-references: EMBL:X52046; NID:g575321; PIDN:CAA36279.1; PID:g575322
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNA
A;Reference number: S16176; MUID:91274355
A;Accession: S16373
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1464 <MET>
A;Cross-references: EMBL:X57983; NID:g50476; PIDN:CAA41048.1; PID:g50477
C;Genetics:
A;Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3;
58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 97
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-154/Domain: propeptide #status predicted <PRO>
F;32-92/Domain: von Willebrand factor type C repeat homology <VWC>
F;155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             70.0%;  Score 233;  DB 2;  Length 1464;
  Best Local Similarity   72.7%;  Pred. No. 3e-12;
  Matches   40;  Conservative    3;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           | || :|: |||| || ||| ||||  |: |||||||| | ||| ||||||||||
Db     530 GGPGIRGMPGSPGGPGNDGKPGPPGSQGESGRPGPPGPSGPRGQPGVMGFPGPKG 584


RESULT  14
I50694
collagen alpha 1(III) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C;Accession: I50694
R;Nah, H.D.; Niu, Z.; Adams, S.L.
J. Biol. Chem. 269, 16443-16448, 1994
A;Title: An alternative transcript of the chick type III collagen gene that does not
A;Reference number: A54041; MUID:94266842
A;Accession: I50694
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-886 <NAH>
A;Cross-references: EMBL:U07973; NID:g520454; PIDN:AAA83407.1; PID:g537432
C;Genetics:
A;Gene: COL3A1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
F;30-90/Domain: von Willebrand factor type C repeat homology <VWC>


  Query Match             69.7%;  Score 232;  DB 2;  Length 886;
  Best Local Similarity   74.5%;  Pred. No. 2.4e-12;
  Matches   41;  Conservative    2;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           |||| :|| | ||||| ||| ||||  |: || ||||| | ||| ||||||||||
Db     530 GLPGMRGLPGIPGSPGSDGKPGPPGNQGEPGRSGPPGPAGPRGQPGVMGFPGPKG 584


RESULT  15
CGBO7S
```

collagen alpha 1(III) chain - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 07-May-1999
C;Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946
R;Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
A;Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen
A;Reference number: A02862; MUID:80026026
A;Accession: A02862
A;Molecule type: protein
A;Residues: 1-242 <FIE>
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
A;Title: The covalent structure of calf skin type III collagen. II. The amino acid seque
A;Reference number: A38001; MUID:80026027
A;Accession: A38001
A;Molecule type: protein
A;Residues: 243-422 <DEW1>
R;Bentz, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
A;Title: The covalent structure of calf skin type III collagen. III. The amino acid sequ
A;Reference number: A38002; MUID:80026028
A;Accession: A38002
A;Molecule type: protein
A;Residues: 423-571 <BEN>
R;Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
A;Title: The covalent structure of calf skin type III collagen. IV. The amino acid seque
A;Reference number: A38003; MUID:80026029
A;Accession: A38003
A;Molecule type: protein
A;Residues: 572-808 <LAN>
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
A;Title: The covalent structure of calf skin type III collagen. V. The amino acid sequen
A;Reference number: A38004; MUID:80026030
A;Accession: A38004
A;Molecule type: protein
A;Residues: 809-947 <DEW2>
R;Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
A;Title: The covalent structure of calf skin type III collagen. VI. The amino acid seque
A;Reference number: A38005; MUID:80026031
A;Accession: A38005
A;Molecule type: protein
A;Residues: 948-1049 <ALL>
A;Experimental source: skin
R;Henkel, W.
Biochem. J. 318, 497-503, 1996
A;Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
A;Reference number: S71946; MUID:96404897
A;Accession: S71946
A;Molecule type: protein
A;Residues: 87-106;1017-1029;1037-1049 <HEN>
C;Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are h
C;Comment: The type III collagen molecule is a trimer of identical chains, linked to eac
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
F;1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>
F;1-14/Region: amino-terminal nonhelical telopeptide
F;15-1040/Region: helical
F;587-589/Region: cell attachment (R-G-D) motif
F;752-754/Region: cell attachment (R-G-D) motif
F;875-877/Region: cell attachment (R-G-D) motif
F;878-880/Region: cell attachment (R-G-D) motif
F;935-937/Region: cell attachment (R-G-D) motif
F;1041-1049/Region: carboxyl-terminal nonhelical telopeptide
F;95,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental
F;107,950/Modified site: allysine (Lys) #status predicted
F;107/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;1040,1041/Disulfide bonds: interchain #status predicted

Query Match 68.5%; Score 228; DB 1; Length 1049;
Best Local Similarity 70.9%; Pred. No. 5.9e-12;
Matches 39; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy 3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
Db 375 GGPGLRGIPGSPGGPGSNGKPGPPGSQGETGRPGPPGSPGPRGQPGVMGFPGPKG 429

Search completed: January 28, 2002, 07:47:26
Job time: 31 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:49:43 ; Search time 38.34 Seconds
(without alignments)
193.201 Million cell updates/sec

Title: US-09-710-239-29
Perfect score: 580
Sequence: 1 RGDKGETGEQGDRGIKGHRG.....DAGPVGPPGPPGPPGPPGPP 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	580	100.0	100	22	AAE02715	Recombinant human
2	580	100.0	100	22	AAB68069	Amino acid sequenc
3	580	100.0	200	22	AAE02714	Recombinant human
4	580	100.0	200	22	AAB68068	Amino acid sequenc
5	580	100.0	219	21	AAY84553	A C-terminal fragm
6	580	100.0	219	21	AAY84555	A C-terminal fragm
7	580	100.0	219	21	AAY84402	C-terminal 219 ami
8	580	100.0	333	22	AAE02713	Recombinant human
9	580	100.0	333	22	AAB68067	Amino acid sequenc
10	580	100.0	441	22	AAG75593	Human colon cancer
11	580	100.0	449	21	AAB43439	Human cancer assoc
12	580	100.0	510	22	AAE02712	Recombinant human
13	580	100.0	510	22	AAB68066	Amino acid sequenc
14	580	100.0	662	22	AAE02718	Human alpha1 (I) t
15	580	100.0	662	22	AAB68072	Amino acid sequenc
16	580	100.0	1057	21	AAY84541	Amino acid sequenc
17	580	100.0	1057	21	AAY84544	A human collagen 1
18	580	100.0	1058	21	AAY84403	Amino acid sequenc
19	580	100.0	1107	17	AAR89472	Collagen/decorin(a
20	580	100.0	1107	21	AAY84540	Amino acid sequenc
21	580	100.0	1169	17	AAR89469	Collagen/BMP-2B fu
22	580	100.0	1169	21	AAY84537	Amino acid sequenc
23	580	100.0	1171	17	AAR89470	Collagen/TGF-beta-
24	580	100.0	1171	21	AAY84538	A chimeric collage
25	580	100.0	1341	16	AAR71701	Collagen alpha 1 (
26	580	100.0	1341	21	AAY96122	Collagen type I al
27	580	100.0	1388	17	AAR89471	Collagen/decorin f
28	580	100.0	1388	21	AAY84539	Amino acid sequenc
29	580	100.0	1449	22	AAE02535	Porcine alpha1(I)
30	580	100.0	1464	19	AAW68485	Human recombinant
31	580	100.0	1464	22	AAU14136	Human novel protei
32	580	100.0	1464	22	AAB82454	Human pro-alpha-1
33	573	98.8	1463	22	AAE02532	Bovine alpha1(I) c
34	472	81.4	1442	16	AAR79480	Rat type II collag
35	461	79.5	1418	16	AAR71703	Collagen alpha 1 (
36	461	79.5	1418	21	AAY96124	Collagen type II a
37	461	79.5	1487	19	AAW61562	Human type II coll
38	453	78.1	1418	15	AAR59751	Type II collagen.
39	453	78.1	1418	22	AAB35624	Human type II coll
40	450	77.6	936	22	AAB70107	Gelatin protein.
41	392	67.6	1196	13	AAR28916	Type III procollag
42	391	67.4	1466	22	AAE02537	Porcine alpha1(III
43	390	67.2	1466	22	AAE02533	Bovine alpha1(III)
44	390	67.2	1466	22	AAE02534	Bovine alpha1(III)
45	384	66.2	1078	16	AAR71704	Collagen alpha 1 (

ALIGNMENTS

RESULT 1
AAE02715
ID AAE02715 standard; Protein; 100 AA.
XX
AC AAE02715;
XX
DT 06-AUG-2001 (first entry)
XX
DE Recombinant human gelatin #4.
XX
KW Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW encapsulant; film-forming agent; moisturising agent; thickening agent;
KW gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW plasma expander; colloidal volume replacement material; graft coating;
KW medical sponge; medical plug; micro-carrier; edible composition;
KW protein supplement; fat substitute; nutritional supplement; cell culture;
KW edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW cartilage degeneration; joint flexibility; food industry; beverage.
XX
OS Homo sapiens.
XX
PN WO200134646-A2.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-US30791.
XX
PR 12-NOV-1999; 99US-0165114.
PR 15-MAY-2000; 2000US-0204437.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
XX

DR WPI; 2001-329072/34.
XX
PT Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT prepared recombinantly -
XX
PS Disclosure; Page 133-134; 137pp; English.
XX
CC The patent discloses recombinant human gelatin which is useful
CC in various compositions including binding agents, encapsulants,
CC stabilising agents, film-forming agents, moisturising agents,
CC emulsifiers, thickening agents, gelling agents, colloidal agents,
CC adhesive agents, pharmaceutical compositions, hard gel capsules,
CC soft gel capsules, plasma expander, colloidal volume replacement
CC materials, graft coatings, medical sponges, medical plugs,
CC pharmaceutical stabilisers, micro-carriers, edible compositions,
CC protein supplements, fat substitutes, nutritional supplements,
CC edible coatings, photographic compositions, cosmetic compositions,
CC industrial composition, cell culture compositions and compositions
CC for use in the laboratory. Pharmaceutical compositions comprising
CC recombinant gelatin are used as vaccines. They are also used to
CC treat various joint conditions such as arthritis, athrosis and
CC other conditions related to the degeneration of cartilage and joint
CC flexibility. Recombinant gelatin is also used in food and beverage
CC industries. The present sequence is a recombinant human gelatin.
XX
SQ Sequence 100 AA;

```
  Query Match             100.0%;  Score 580;  DB 22;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 60

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||||
Db      61 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 100
```

RESULT 2
AAB68069
ID AAB68069 standard; Protein; 100 AA.
XX
AC AAB68069;
XX
DT 09-JUL-2001 (first entry)
XX
DE Amino acid sequence of a recombinant human gelatin.
XX
KW Human; gelatin; vaccine; anaphylactic reaction.
XX
OS Homo sapiens.
XX
PN WO200134801-A2.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-US30843.
XX
PR 12-NOV-1999; 99US-0165114.
PR 15-MAY-2000; 2000US-0204437.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
XX
DR WPI; 2001-308784/32.
XX
PT Vaccine formulations (I) comprising recombinant human gelatin, useful
PT for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT and cholera, the gelatin is non-immunogenic and confers stability at
PT ambient temperatures -
XX
PS Claim 11; Page 126-127; 130pp; English.
XX
CC The present sequence represents a human recombinant gelatin polypeptide.
CC The recombinant gelatin polypeptide is used to produce vaccine
CC formulations of the invention. The recombinant human gelatin is
CC non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC stability at ambient temperatures. The vaccine formulation comprises a
CC vaccine formulated for the prevention of a disease selected from vaccinia
CC virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC haemophillus influenzae meningitis, rabies, cholera, Japanese
CC encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ Sequence 100 AA;

```
  Query Match             100.0%;  Score 580;  DB 22;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 60

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||||
Db      61 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 100
```

RESULT 3
AAE02714
ID AAE02714 standard; Protein; 200 AA.
XX
AC AAE02714;
XX
DT 06-AUG-2001 (first entry)
XX
DE Recombinant human gelatin #3.
XX
KW Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW encapsulant; film-forming agent; moisturising agent; thickening agent;
KW gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW plasma expander; colloidal volume replacement material; graft coating;
KW medical sponge; medical plug; micro-carrier; edible composition;
KW protein supplement; fat substitute; nutritional supplement; cell culture;
KW edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW cartilage degeneration; joint flexibility; food industry; beverage.
XX
OS Homo sapiens.
XX
PN WO200134646-A2.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-US30791.
XX
PR 12-NOV-1999; 99US-0165114.
PR 15-MAY-2000; 2000US-0204437.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
XX
DR WPI; 2001-329072/34.
XX
PT Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT prepared recombinantly -

```
XX
PS   Example 1; Page 133; 137pp; English.
XX
CC   The patent discloses recombinant human gelatin which is useful
CC   in various compositions including binding agents, encapsulants,
CC   stabilising agents, film-forming agents, moisturising agents,
CC   emulsifiers, thickening agents, gelling agents, colloidal agents,
CC   adhesive agents, pharmaceutical compositions, hard gel capsules,
CC   soft gel capsules, plasma expander, colloidal volume replacement
CC   materials, graft coatings, medical sponges, medical plugs,
CC   pharmaceutical stabilisers, micro-carriers, edible compositions,
CC   protein supplements, fat substitutes, nutritional supplements,
CC   edible coatings, photographic compositions, cosmetic compositions,
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is a recombinant human gelatin.
XX
SQ   Sequence   200 AA;


  Query Match            100.0%;  Score 580;  DB 22;  Length 200;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 160

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     161 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 200


RESULT   4
AAB68068
ID   AAB68068 standard; Protein; 200 AA.
XX
AC   AAB68068;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 126; 130pp; English.
XX
CC   The present sequence represents a human recombinant gelatin polypeptide.
CC   The recombinant gelatin polypeptide is used to produce vaccine
CC   formulations of the invention. The recombinant human gelatin is
CC   non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC   stability at ambient temperatures. The vaccine formulation comprises a
CC   vaccine formulated for the prevention of a disease selected from vaccinia
CC   virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC   diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC   (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC   haemophillus influenzae meningitis, rabies, cholera, Japanese
CC   encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC   adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC   respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC   herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ   Sequence   200 AA;


  Query Match            100.0%;  Score 580;  DB 22;  Length 200;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 160

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     161 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 200


RESULT   5
AAY84553
ID   AAY84553 standard; Protein; 219 AA.
XX
AC   AAY84553;
XX
DT   25-JUL-2000  (first entry)
XX
DE   A C-terminal fragment of human collagen type 1 (alpha1).
XX
KW   Extracellular matrix protein; self aggregation; hydroxylated proline;
KW   trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW   collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS   Homo sapiens.
XX
PN   EP992586-A2.
XX
PD   12-APR-2000.
XX
PF   07-OCT-1999;   99EP-0119184.
XX
PR   09-OCT-1998;   98US-0169768.
XX
PA   (USSU ) US SURGICAL CORP.
XX
PI   Gruskin EA,  Buechter DD,  Zhang G,  Connolly K;
XX
DR   WPI; 2000-259138/23.
DR   N-PSDB; AAA12518.
XX
PT   Production of extracellular matrix proteins containing
PT   4-trans-hydroxyproline results in native self aggregating proteins,
PT   useful on medical implants -
XX
PS   Claim 16; Fig 72; 260pp; English.
XX
CC   The specification describes a method for producing an extracellular
CC   matrix protein or its fragment. The extracellular matrix protein is
CC   capable of self aggregating in a cell which does not ordinarily
CC   hydroxylated prolines. The method comprises optimising a nucleic acid
CC   sequence for expression in the cell by substitution of codons preferred
```

CC by that cell for naturally occurring codons not preferred by the cell;
CC incorporating the nucleic acid sequence into the cell; and contacting
CC the cell with a hypertonic growth medium containing at least one amino
CC acid, selected from the group consisting of trans-4-hydroxyproline and
CC 3-hydroxyproline to allow at least one of the amino acids to be
CC assimilated into the cell and incorporated into the extracellular matrix
CC protein. The method may be used to make host cells assimilate and
CC incorporate trans-4-hydroxyproline into proteins. This is especially
CC useful in the recombinant production of proteins such as collagen,
CC fibrinogen and fibronectin whose ability to self aggregate and produce
CC functional proteins depends on the post translational hydroxylation of
CC proline. The method is also useful in studying the structure and function
CC of polypeptides which do not normally contain trans-4-hydroxyproline.
CC The present sequence represents a C-terminal fragment of human collagen
CC type 1 (alpha1), with optimised codon usage, designated D4.
XX
SQ Sequence 219 AA;

Query Match 100.0%; Score 580; DB 21; Length 219;
Best Local Similarity 100.0%; Pred. No. 3.1e-35;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
Db 94 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 153

Qy 61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
Db 154 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 193

RESULT 6
AAY84555
ID AAY84555 standard; Protein; 219 AA.
XX
AC AAY84555;
XX
DT 25-JUL-2000 (first entry)
XX
DE A C-terminal fragment of human collagen type 1 (alpha2).
XX
KW Extracellular matrix protein; self aggregation; hydroxylated proline;
KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS Homo sapiens.
XX
PN EP992586-A2.
XX
PD 12-APR-2000.
XX
PF 07-OCT-1999; 99EP-0119184.
XX
PR 09-OCT-1998; 98US-0169768.
XX
PA (USSU) US SURGICAL CORP.
XX
PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX
DR WPI; 2000-259138/23.
XX
PT Production of extracellular matrix proteins containing
PT 4-trans-hydroxyproline results in native self aggregating proteins,
PT useful on medical implants -
XX
PS Claim 10; Fig 80; 260pp; English.
XX
CC The specification describes a method for producing an extracellular
CC matrix protein or its fragment. The extracellular matrix protein is
CC capable of self aggregating in a cell which does not ordinarily
CC hydroxylated prolines. The method comprises optimising a nucleic acid
CC sequence for expression in the cell by substitution of codons preferred

CC by that cell for naturally occurring codons not preferred by the cell;
CC incorporating the nucleic acid sequence into the cell; and contacting
CC the cell with a hypertonic growth medium containing at least one amino
CC acid, selected from the group consisting of trans-4-hydroxyproline and
CC 3-hydroxyproline to allow at least one of the amino acids to be
CC assimilated into the cell and incorporated into the extracellular matrix
CC protein. The method may be used to make host cells assimilate and
CC incorporate trans-4-hydroxyproline into proteins. This is especially
CC useful in the recombinant production of proteins such as collagen,
CC fibrinogen and fibronectin whose ability to self aggregate and produce
CC functional proteins depends on the post translational hydroxylation of
CC proline. The method is also useful in studying the structure and function
CC of polypeptides which do not normally contain trans-4-hydroxyproline.
CC The present sequence represents a C-terminal fragment of human collagen
CC type 1 (alpha2), with optimised codon usage, designated D4.
XX
SQ Sequence 219 AA;

Query Match 100.0%; Score 580; DB 21; Length 219;
Best Local Similarity 100.0%; Pred. No. 3.1e-35;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
Db 94 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 153

Qy 61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
Db 154 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 193

RESULT 7
AAY84402
ID AAY84402 standard; Protein; 219 AA.
XX
AC AAY84402;
XX
DT 12-JUL-2000 (first entry)
XX
DE C-terminal 219 amino acids of human alpha1 collagen.
XX
KW Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
KW collagen; mussel adhesive protein; bioadhesive.
XX
OS Homo sapiens.
XX
PN WO200014201-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-US20462.
XX
PR 09-SEP-1998; 98US-0099652.
XX
PA (USSU) US SURGICAL CORP.
PA (PAOL/) PAOLELLA D N.
PA (GRUS/) GRUSKIN E A.
PA (BUEC/) BUECHTER D D.
XX
PI Paolella DN, Gruskin EA, Buechter DD;
XX
DR WPI; 2000-271051/23.
DR N-PSDB; AAZ99842.
XX
PT Incorporating non-natural amino acid into polypeptide, useful e.g. for
PT production of bioadhesives, by epoxidation or substitution of
PT dehydroproline residues -
XX
PS Disclosure; Fig 4; 66pp; English.
XX
CC The present sequence represents the C-terminal 219 amino acids of
CC the human alpha1 collagen protein. Peptides derived from the protien

```
CC   were used to demonstrate incorporation of 3,4-dehydro-L-proline into
CC   the peptide, using the method of the invention. The specification
CC   describes a method for the incorporation of non-natural amino acid
CC   into a polypeptide. The method comprises reacting at least one
CC   3,4-dehydroproline residue in the polypeptide with an epoxidation
CC   reagent from a polypeptide containing at least one 3,4-epoxyproline
CC   residue. The method is used for studying the effects of non-natural
CC   amino acids on structure and function of polypeptides. The method is
CC   also useful for commercial production of collagen or mussel adhesive
CC   proteins (which are useful as bioadhesives), and for incorporating a
CC   wide variety of groups, including therapeutic ligands and biological
CC   probes, into polypeptides.
XX
SQ   Sequence   219 AA;


  Query Match             100.0%;  Score 580;  DB 21;  Length 219;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 153

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           |||||||||||||||||||||||||||||||||||||||||
Db     154 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 193


RESULT   8
AAE02713
ID   AAE02713 standard; Protein; 333 AA.
XX
AC   AAE02713;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Recombinant human gelatin #2.
XX
KW   Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW   encapsulant; film-forming agent; moisturising agent; thickening agent;
KW   gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW   plasma expander; colloidal volume replacement material; graft coating;
KW   medical sponge; medical plug; micro-carrier; edible composition;
KW   protein supplement; fat substitute; nutritional supplement; cell culture;
KW   edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW   cartilage degeneration; joint flexibility; food industry; beverage.
XX
OS   Homo sapiens.
XX
PN   WO200134646-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30791.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-329072/34.
XX
PT   Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT   prepared recombinantly -
XX
PS   Example 1; Page 132-133; 137pp; English.
XX
CC   The patent discloses recombinant human gelatin which is useful
CC   in various compositions including binding agents, encapsulants,
CC   stabilising agents, film-forming agents, moisturising agents,
CC   emulsifiers, thickening agents, gelling agents, colloidal agents,
CC   adhesive agents, pharmaceutical compositions, hard gel capsules,
CC   soft gel capsules, plasma expander, colloidal volume replacement
CC   materials, graft coatings, medical sponges, medical plugs,
CC   pharmaceutical stabilisers, micro-carriers, edible compositions,
CC   protein supplements, fat substitutes, nutritional supplements,
CC   edible coatings, photographic compositions, cosmetic compositions,
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is a recombinant human gelatin.
XX
SQ   Sequence   333 AA;


  Query Match             100.0%;  Score 580;  DB 22;  Length 333;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     234 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 293

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           |||||||||||||||||||||||||||||||||||||||||
Db     294 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 333


RESULT   9
AAB68067
ID   AAB68067 standard; Protein; 333 AA.
XX
AC   AAB68067;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 125-126; 130pp; English.
XX
CC   The present sequence represents a human recombinant gelatin polypeptide.
CC   The recombinant gelatin polypeptide is used to produce vaccine
CC   formulations of the invention. The recombinant human gelatin is
CC   non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC   stability at ambient temperatures. The vaccine formulation comprises a
CC   vaccine formulated for the prevention of a disease selected from vaccinia
```

```
CC   virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC   diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC   (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC   haemophillus influenzae meningitis, rabies, cholera, Japanese
CC   encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC   adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC   respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC   herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ   Sequence   333 AA;


  Query Match             100.0%;  Score 580;  DB 22;  Length 333;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     234 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 293

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     294 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 333


RESULT  10
AAG75593
ID   AAG75593 standard; Protein; 441 AA.
XX
AC   AAG75593;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen protein SEQ ID NO:6357.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; chromosome 17.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   N-PSDB; AAH34998.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 11; Page 7817-7819; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence   441 AA;


  Query Match             100.0%;  Score 580;  DB 22;  Length 441;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 137

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     138 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 177


RESULT  11
AAB43439
ID   AAB43439 standard; Protein; 449 AA.
XX
AC   AAB43439;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated protein sequence SEQ ID NO:884.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   N-PSDB; AAC77648.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer -
XX
PS   Claim 11; Page 1439-1441; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
```

```
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   449 AA;


  Query Match             100.0%;  Score 580;  DB 21;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 137

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     138 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 177


RESULT  12
AAE02712
ID   AAE02712 standard; Protein; 510 AA.
XX
AC   AAE02712;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Recombinant human gelatin #1.
XX
KW   Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW   encapsulant; film-forming agent; moisturising agent; thickening agent;
KW   gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW   plasma expander; colloidal volume replacement material; graft coating;
KW   medical sponge; medical plug; micro-carrier; edible composition;
KW   protein supplement; fat substitute; nutritional supplement; cell culture;
KW   edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW   cartilage degeneration; joint flexibility; food industry; beverage.
XX
OS   Homo sapiens.
XX
PN   WO200134646-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30791.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-329072/34.
XX
PT   Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT   prepared recombinantly -
XX
PS   Disclosure; Page 130-131; 137pp; English.
XX
CC   The patent discloses recombinant human gelatin which is useful
CC   in various compositions including binding agents, encapsulants,
CC   stabilising agents, film-forming agents, moisturising agents,
CC   emulsifiers, thickening agents, gelling agents, colloidal agents,
CC   adhesive agents, pharmaceutical compositions, hard gel capsules,
CC   soft gel capsules, plasma expander, colloidal volume replacement
CC   materials, graft coatings, medical sponges, medical plugs,
CC   pharmaceutical stabilisers, micro-carriers, edible compositions,
CC   protein supplements, fat substitutes, nutritional supplements,
CC   edible coatings, photographic compositions, cosmetic compositions,
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is a recombinant human gelatin.
XX
SQ   Sequence   510 AA;


  Query Match             100.0%;  Score 580;  DB 22;  Length 510;
  Best Local Similarity   100.0%;  Pred. No. 6e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 470

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     471 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 510


RESULT  13
AAB68066
ID   AAB68066 standard; Protein; 510 AA.
XX
AC   AAB68066;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 123-124; 130pp; English.
XX
CC   The present sequence represents a human recombinant gelatin polypeptide.
CC   The recombinant gelatin polypeptide is used to produce vaccine
CC   formulations of the invention. The recombinant human gelatin is
```

```
CC   non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC   stability at ambient temperatures. The vaccine formulation comprises a
CC   vaccine formulated for the prevention of a disease selected from vaccinia
CC   virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC   diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC   (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC   haemophillus influenzae meningitis, rabies, cholera, Japanese
CC   encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC   adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC   respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC   herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ   Sequence   510 AA;


  Query Match             100.0%;  Score 580;  DB 22;  Length 510;
  Best Local Similarity   100.0%;  Pred. No. 6e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     411 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 470

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     471 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 510


RESULT  14
AAE02718
ID   AAE02718 standard; Protein; 662 AA.
XX
AC   AAE02718;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Human alpha1 (I) type I collagen helical domain (residues 531-1192).
XX
KW   Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
KW   encapsulant; film-forming agent; moisturising agent; thickening agent;
KW   gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
KW   plasma expander; colloidal volume replacement material; graft coating;
KW   medical sponge; medical plug; micro-carrier; edible composition;
KW   protein supplement; fat substitute; nutritional supplement; cell culture;
KW   edible coating; cosmetic; vaccine; therapy; arthritis; athrosis;
KW   cartilage degeneration; joint flexibility; food industry; beverage;
KW   alpha1 (I) type I collagen.
XX
OS   Homo sapiens.
XX
PN   WO200134646-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30791.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-329072/34.
XX
PT   Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
PT   prepared recombinantly -
XX
PS   Claim 21; Page 135-137; 137pp; English.
XX
CC   The patent discloses recombinant human gelatin which is useful
CC   in various compositions including binding agents, encapsulants,
CC   stabilising agents, film-forming agents, moisturising agents,
CC   emulsifiers, thickening agents, gelling agents, colloidal agents,
CC   adhesive agents, pharmaceutical compositions, hard gel capsules,
CC   soft gel capsules, plasma expander, colloidal volume replacement
CC   materials, graft coatings, medical sponges, medical plugs,
CC   pharmaceutical stabilisers, micro-carriers, edible compositions,
CC   protein supplements, fat substitutes, nutritional supplements,
CC   edible coatings, photographic compositions, cosmetic compositions,
CC   industrial composition, cell culture compositions and compositions
CC   for use in the laboratory. Pharmaceutical compositions comprising
CC   recombinant gelatin are used as vaccines. They are also used to
CC   treat various joint conditions such as arthritis, athrosis and
CC   other conditions related to the degeneration of cartilage and joint
CC   flexibility. Recombinant gelatin is also used in food and beverage
CC   industries. The present sequence is human alpha1 (I) type I collagen
CC   helical domain (residues 531-1192). This sequence is a recombinant
CC   gelatin.
XX
SQ   Sequence   662 AA;


  Query Match             100.0%;  Score 580;  DB 22;  Length 662;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-35;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     563 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 622

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     623 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 662


RESULT  15
AAB68072
ID   AAB68072 standard; Protein; 662 AA.
XX
AC   AAB68072;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Amino acid sequence of a recombinant human gelatin.
XX
KW   Human; gelatin; vaccine; anaphylactic reaction.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 53
FT                   /note= "this residue is given as unknown as it is
FT                         illegible in the specification"
XX
PN   WO200134801-A2.
XX
PD   17-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-US30843.
XX
PR   12-NOV-1999;   99US-0165114.
PR   15-MAY-2000; 2000US-0204437.
XX
PA   (FIBR-) FIBROGEN INC.
XX
PI   Chang RC,  Kivirikko KI,  Neff TB,  Olsen DR,  Polarek JW;
XX
DR   WPI; 2001-308784/32.
XX
PT   Vaccine formulations (I) comprising recombinant human gelatin, useful
PT   for vaccinating against e.g. mumps, measles, rubella, tetanus, rabies
PT   and cholera, the gelatin is non-immunogenic and confers stability at
PT   ambient temperatures -
XX
PS   Claim 11; Page 128-130; 130pp; English.
```

XX
CC The present sequence represents a human recombinant gelatin polypeptide.
CC The recombinant gelatin polypeptide is used to produce vaccine
CC formulations of the invention. The recombinant human gelatin is
CC non-immunogenic (therefore reducing anaphylactic reactions) and confers
CC stability at ambient temperatures. The vaccine formulation comprises a
CC vaccine formulated for the prevention of a disease selected from vaccinia
CC virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
CC diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
CC (whooping cough), Bacille Calmette-Geurin (BCG, tuberculosis),
CC haemophillus influenzae meningitis, rabies, cholera, Japanese
CC encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
CC adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
CC respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
CC herpes virus (Marek's disease), influenza and/or anthrax.
XX
SQ Sequence 662 AA;

Query Match 100.0%; Score 580; DB 22; Length 662;
Best Local Similarity 100.0%; Pred. No. 7.4e-35;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     563 rgdkgetgeqgdrgikghrgfsglqgppgppgspgeqgpsgasgpagprgppgsagapgk 622

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||
Db     623 dglnglpgpigppgprgrtgdagpvgppgppgppgppgpp 662
```

Search completed: January 28, 2002, 07:49:44
Job time: 169 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:46:55 ; Search time 15.36 Seconds
(without alignments)
140.835 Million cell updates/sec

Title: US-09-710-239-18
Perfect score: 333
Sequence: 1 EAGLPGAKGLTGSPGSPGPD.....PPGARGQAGVMGFPGPKGAA 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	333	100.0	1460	1	CA11_CANFA	Q9xsj7 canis famil
2	333	100.0	1464	1	CA11_HUMAN	P02452 homo sapien
3	325	97.6	1453	1	CA11_CHICK	P02457 gallus gall
4	321	96.4	1453	1	CA11_MOUSE	P11087 mus musculu
5	317	95.2	671	1	CA11_RAT	P02454 rattus norv
6	264	79.3	747	1	CA12_BOVIN	P02459 bos taurus
7	256	76.9	1418	1	CA12_HUMAN	P02458 homo sapien
8	256	76.9	1459	1	CA12_MOUSE	P28481 mus musculu
9	234	70.3	1466	1	CA13_HUMAN	P02461 homo sapien
10	233	70.0	1464	1	CA13_MOUSE	P08121 mus musculu
11	232	69.7	1262	1	CA13_CHICK	P12105 gallus gall
12	228	68.5	1049	1	CA13_BOVIN	P04258 bos taurus
13	224	67.3	1496	1	CA25_HUMAN	P05997 homo sapien
14	216	64.9	1356	1	CA21_ONCMY	O93484 oncorhynchu
15	204	61.3	1372	1	CA21_MOUSE	Q01149 mus musculu
16	202	60.7	1372	1	CA21_RAT	P02466 rattus norv
17	201	60.4	1366	1	CA21_HUMAN	P08123 homo sapien
18	197	59.2	1364	1	CA21_BOVIN	P02465 bos taurus
19	195	58.6	1366	1	CA21_CANFA	O46392 canis famil
20	193	58.0	1355	1	CA21_RANCA	O42350 rana catesb
21	191	57.4	1763	1	CA24_ASCSU	P27393 ascaris suu
22	183	55.0	296	1	CC01_CAEEL	P08124 caenorhabdi
23	183	55.0	1685	1	CA54_HUMAN	P29400 homo sapien
24	181.5	54.5	1669	1	CA14_MOUSE	P02463 mus musculu
25	180	54.1	675	1	CA39_CHICK	P32017 gallus gall
26	180	54.1	1690	1	CA44_HUMAN	P53420 homo sapien
27	179.5	53.9	674	1	CA1A_BOVIN	P23206 bos taurus
28	179.5	53.9	680	1	CA1A_HUMAN	Q03692 homo sapien
29	176.5	53.0	1669	1	CA14_HUMAN	P02462 homo sapien
30	176	52.9	325	1	CA19_RAT	P20850 rattus norv
31	175.5	52.7	754	1	CA54_CANFA	Q28247 canis famil
32	174	52.3	295	1	CAC3_HAECO	P16253 haemonchus
33	173.5	52.1	636	1	CA13_RAT	P13941 rattus norv
34	173.5	52.1	1527	1	CA1H_MOUSE	P39061 mus musculu
35	173.5	52.1	1838	1	CA15_HUMAN	P20908 homo sapien
36	173	52.0	2944	1	CA17_HUMAN	Q02388 homo sapien
37	172.5	51.8	779	1	CA11_BOVIN	P02453 bos taurus
38	171.5	51.5	1670	1	CA34_HUMAN	Q01955 homo sapien
39	171	51.4	931	1	CA19_HUMAN	P20849 homo sapien
40	169.5	50.9	1028	1	CA16_HUMAN	P12109 homo sapien
41	169	50.8	316	1	CC12_CAEEL	P20630 caenorhabdi
42	169	50.8	316	1	CC13_CAEEL	P20631 caenorhabdi
43	169	50.8	921	1	CA19_MOUSE	Q05722 mus musculu
44	169	50.8	1804	1	CA1B_MOUSE	Q61245 mus musculu
45	168.5	50.6	471	1	CA34_BOVIN	Q28084 bos taurus

ALIGNMENTS

RESULT 1
CA11_CANFA
ID CA11_CANFA STANDARD; PRT; 1460 AA.
AC Q9XSJ7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN COL1A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC ---

CC ---
DR EMBL; AF153062; AAD34619.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22 BY SIMILARITY.
FT PROPEP 23 157 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 158 1214 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1215 1460 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 34 92 VWFC.
FT DOMAIN 158 174 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 175 1188 TRIPLE-HELICAL REGION.
FT DOMAIN 1189 1214 NONHELICAL REGION (C-TERMINAL).
FT SITE 741 743 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).

```
FT   CARBOHYD   1361   1361       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1460 AA;  138762 MW;  58E3674D2B570697 CRC64;


  Query Match             100.0%;  Score 333;  DB 1;  Length 1460;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-19;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    527 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 585


RESULT   2
CA11_HUMAN
ID   CA11_HUMAN     STANDARD;      PRT;  1464 AA.
AC   P02452; Q15176; Q14037;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-472 FROM N.A.
RX   MEDLINE=89025644; PubMed=3178743;
RA   Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA   Jaenisch R., Prockup D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT   chain of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [2]
RP   SEQUENCE OF 1-181 FROM N.A.
RX   MEDLINE=84270697; PubMed=6462220;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA   Myers J., Williams C., Ramirez F.;
RT   "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT   conservation of a pattern of introns and exons.";
RL   Nature 310:337-340(1984).
RN   [3]
RP   SEQUENCE OF 162-301.
RC   TISSUE=Skin;
RX   MEDLINE=71038625; PubMed=5529814;
RA   Click E.M., Bornstein P.;
RT   "Isolation and characterization of the cyanogen bromide peptides from
RT   the alpha 1 and alpha 2 chains of human skin collagen.";
RL   Biochemistry 9:4699-4706(1970).
RN   [4]
RP   SEQUENCE OF 263-268.
RC   TISSUE=Skin;
RX   MEDLINE=71001508; PubMed=4319110;
RA   Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT   "A comparative study of glycopeptides derived from selected
RT   vertebrate collagens. A possible role of the carbohydrate in fibril
RT   formation.";
RL   J. Biol. Chem. 245:5042-5048(1970).
RN   [5]
RP   SEQUENCE OF 425-1464 FROM N.A.
RX   MEDLINE=84080385; PubMed=6689127;
RA   Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA   Prockop D.J.;
RT   "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT   pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT   of structures that are conserved during evolution.";
RL   Biochemistry 22:5213-5223(1983).
RN   [6]
RP   SEQUENCE OF 1229-1454 FROM N.A.
RC   TISSUE=Bone;
RX   MEDLINE=88124208; PubMed=3340531;
RA   Maekelae J.K., Raassina M., Virta A., Vuorio E.;
RT   "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT   domain.";
RL   Nucleic Acids Res. 16:349-349(1988).
RN   [7]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=88097389; PubMed=3480516;
RA   Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT   "Regulatory elements in the first intron contribute to
RT   transcriptional control of the human alpha 1(I) collagen gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN   [8]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=85130970; PubMed=2857713;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT   Promoter structure, AluI repeats, and polymorphic transcripts.";
RL   J. Biol. Chem. 260:2315-2320(1985).
RN   [9]
RP   SEQUENCE OF 1-44 FROM N.A.
RX   MEDLINE=88033098; PubMed=2822714;
RA   Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA   de Wet W.J.;
RT   "DNA sequences in the first intron of the human pro-alpha 1(I)
RT   collagen gene enhance transcription.";
RL   J. Biol. Chem. 262:15151-15157(1987).
RN   [10]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=91184577; PubMed=2010058;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in collagen genes: causes of rare and some common diseases
RT   in humans.";
RL   FASEB J. 5:2052-2060(1991).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [12]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE=91374476; PubMed=1895312;
RA   Byers P.H., Wallis G.A., Willing M.C.;
RT   "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL   J. Med. Genet. 28:433-442(1991).
RN   [13]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE=97169389; PubMed=9016532;
RA   Dalgleish R.;
RT   "The human type I collagen mutation database.";
RL   Nucleic Acids Res. 25:181-187(1997).
RN   [14]
RP   VARIANT OI-II CYS-1166.
RX   MEDLINE=86287390; PubMed=3016737;
RA   Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT   "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT   change in one human pro alpha 1(I) collagen allele.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN   [15]
RP   VARIANT OI-II ARG-569.
RX   MEDLINE=87222295; PubMed=3108247;
RA   Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT   "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT   arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT   collagen.";
RL   J. Biol. Chem. 262:7021-7027(1987).
RN   [16]
RP   VARIANT OI-II CYS-926.
RX   MEDLINE=88033031; PubMed=3667599;
RA   Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT   "A point mutation in a type I procollagen gene converts glycine 748
RT   of the alpha 1 chain to cysteine and destabilizes the triple helix in
```

```
RT   a lethal variant of osteogenesis imperfecta.";
RL   J. Biol. Chem. 262:14737-14744(1987).
RN   [17]
RP   VARIANT OI-II ARG-842.
RX   MEDLINE=88298828; PubMed=3403550;
RA   Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT   "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RT   chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT   the peptide defect by in vitro expression of the mutant cDNA.";
RL   J. Biol. Chem. 263:11627-11630(1988).
RN   [18]
RP   VARIANT OI CYS-1195.
RX   MEDLINE=89218628; PubMed=3244312;
RA   Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT   "A cysteine for glycine substitution at position 1017 in an alpha
RT   1(I) chain of type I collagen in a patient with mild dominantly
RT   inherited osteogenesis imperfecta.";
RL   Mol. Biol. Med. 5:197-207(1988).
RN   [19]
RP   VARIANT OI-II VAL-434.
RX   MEDLINE=89255493; PubMed=2470760;
RA   Patterson E., Smiley E., Bonadio J.;
RT   "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT   mutation.";
RL   J. Biol. Chem. 264:10083-10087(1989).
RN   [20]
RP   VARIANT OI-IV SER-1010.
RX   MEDLINE=89308591; PubMed=2745420;
RA   Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT   "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT   one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL   J. Biol. Chem. 264:11893-11900(1989).
RN   [21]
RP   VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX   MEDLINE=89380165; PubMed=2777764;
RA   Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT   "Characterization of point mutations in the collagen COL1A1 and
RT   COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL   J. Biol. Chem. 264:15809-15812(1989).
RN   [22]
RP   VARIANT OI SER-1022.
RX   MEDLINE=90062068; PubMed=2511192;
RA   Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT   "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT   variant of osteogenesis imperfecta minimally destabilizes the triple
RT   helix of type I procollagen. The effects of glycine substitutions on
RT   thermal stability are either position of amino acid specific.";
RL   J. Biol. Chem. 264:19694-19699(1989).
RN   [23]
RP   VARIANT OI-II CYS-1082.
RX   MEDLINE=89109573; PubMed=2913053;
RA   Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT   "A lethal variant of osteogenesis imperfecta has a single base
RT   mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT   chain of type I procollagen. The asymptomatic mother has an
RT   unidentified mutation producing an overmodified and unstable type I
RT   procollagen.";
RL   J. Clin. Invest. 83:574-584(1989).
RN   [24]
RP   VARIANT OI CYS-272; CYS-704 AND CYS-896.
RX   MEDLINE=90009313; PubMed=2794057;
RA   Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
RA   Weiss L., Graham J.M., Byers P.H.;
RT   "Osteogenesis imperfecta. The position of substitution for glycine by
RT   cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT   type I collagen determines the clinical phenotype.";
RL   J. Clin. Invest. 84:1206-1214(1989).
RN   [25]
RP   VARIANT OI-II CYS-422.

  Query Match            100.0%;  Score 333;  DB 1;  Length 1464;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-19;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     531 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 589


RESULT   3
CA11_CHICK
ID   CA11_CHICK     STANDARD;      PRT;  1453 AA.
AC   P02457;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-153 FROM N.A.
RX   MEDLINE=88056316; PubMed=3678834;
RA   Finer M.H., Boedtker H., Doty P.;
RT   "Construction and characterization of cDNA clones encoding the 5' end
RT   of the chicken pro alpha 1(I) collagen mRNA.";
RL   Gene 56:71-78(1987).
RN   [2]
RP   SEQUENCE OF 1-144 FROM N.A.
RX   MEDLINE=88007542; PubMed=2820966;
RA   Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT   "Unusual DNA sequences located within the promoter region and the
RT   first intron of the chicken pro-alpha 1(I) collagen gene.";
RL   J. Biol. Chem. 262:13323-13332(1987).
RN   [3]
RP   SEQUENCE OF 152-1187.
RX   MEDLINE=82231995; PubMed=7093229;
RA   Highberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA   Kang A.H., Gross J.;
RT   "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the
RT   complete primary structure of the helical portion of the chick skin
RT   collagen alpha 1(I) chain.";
RL   Biochemistry 21:2048-2055(1982).
RN   [4]
RP   SEQUENCE OF 1200-1205.
RX   MEDLINE=72243016; PubMed=5047697;
RA   Eyre D.R., Glimcher M.J.;
RT   "Evidence for a previously undetected sequence at the carboxyterminus
RT   of the alpha 1 chain of chicken bone collagen.";
RL   Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN   [5]
RP   SEQUENCE OF 981-1453 FROM N.A.
RX   MEDLINE=81160715; PubMed=6927845;
RA   Fuller F., Boedtker H.;
RT   "Sequence determination and analysis of the 3' region of chicken pro-
RT   alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT   including the carboxy-terminal propeptide sequences.";
RL   Biochemistry 20:996-1006(1981).
RN   [6]
RP   SEQUENCE OF 1311-1453 FROM N.A.
RX   MEDLINE=80134546; PubMed=6987088;
RA   Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA   Pastan I., Decrombrugghe B., Fietzek P.P., Olsen B.R.;
RT   "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT   carboxyl end of pro alpha 1(I)-chains.";
RL   FEBS Lett. 111:61-65(1980).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
```

```
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17839; AAA48704.1; -.
DR   EMBL; M17838; AAA48704.1; JOINED.
DR   EMBL; V00401; CAA23695.1; -.
DR   EMBL; M10571; AAA48671.1; ALT_SEQ.
DR   EMBL; M17607; AAA48672.1; -.
DR   PIR; A02857; CGCH1S.
DR   PIR; A27179; A27179.
DR   PIR; A29367; A29367.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL       1     22
FT   PROPEP      23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN      152   1205       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP    1206   1453       C-TERMINAL PROPEPTIDE.
FT   DOMAIN      31     89       VWFC.
FT   MOD_RES    152    152       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES    254    254       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    851    851       HYDROXYLATION (POTENTIAL).
FT   MOD_RES   1081   1081       HYDROXYLATION (POTENTIAL).
FT   MOD_RES   1097   1097       HYDROXYLATION (POTENTIAL).
FT   MOD_RES   1153   1153       HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE
FT                               ONLY HYDROXYLATED PRO IN POSITION X (IN
FT                               THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
FT   CONFLICT  1187   1187       F -> L (IN REF. 5).
FT   CONFLICT  1441   1441       Q -> H (IN REF. 6).
SQ   SEQUENCE   1453 AA;  137789 MW;  3BC6152134271F4D CRC64;


  Query Match             97.6%;  Score 325;  DB 1;  Length 1453;
  Best Local Similarity   98.3%;  Pred. No. 6.2e-19;
  Matches   58;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     520 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGAA 578


RESULT   4
CA11_MOUSE
ID   CA11_MOUSE     STANDARD;      PRT;  1453 AA.
AC   P11087; Q60635;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1 OR COLA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N;
RX   MEDLINE=96033240; PubMed=8535610;
RA   Li S.W., Khillan J., Prockop D.J.;
RT   "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT   of type I procollagen.";
RL   Matrix Biol. 14:593-595(1995).
RN   [2]
RP   SEQUENCE OF 518-1128 FROM N.A.
RX   MEDLINE=86137403; PubMed=3841523;
RA   French B.T., Lee W.-H., Maul G.G.;
RT   "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT   collagen protein.";
RL   Gene 39:311-312(1985).
RN   [3]
RP   SEQUENCE OF 735-1130 FROM N.A.
RX   MEDLINE=83141374; PubMed=6298597;
RA   Monson J.M., Friedman J., McCarthy B.J.;
RT   "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
RT   evidence for a mouse B1 element within the gene.";
RL   Mol. Cell. Biol. 2:1362-1371(1982).
RN   [4]
RP   SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX   MEDLINE=83157109; PubMed=6219867;
RA   Monson J.M., McCarthy B.J.;
RT   "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT   evidence for insertions or deletions in gene coding sequences.";
RL   DNA 1:59-69(1981).
RN   [5]
RP   SEQUENCE OF 1442-1453 FROM N.A.
RX   MEDLINE=88124276; PubMed=3340560;
RA   Mooslehner K., Harbers K.;
RT   "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT   of the 3'-untranslated region.";
RL   Nucleic Acids Res. 16:773-773(1988).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08020; AAA88912.1; -.
DR   EMBL; X15896; CAA33904.1; -.
DR   EMBL; M14423; AAA37333.1; -.
DR   EMBL; M17491; AAA37334.1; -.
DR   EMBL; X06753; CAA29927.1; -.
DR   EMBL; K03036; AAA37332.1; -.
DR   EMBL; K03029; AAA37332.1; JOINED.
DR   EMBL; K03030; AAA37332.1; JOINED.
DR   EMBL; K03031; AAA37332.1; JOINED.
DR   EMBL; K03032; AAA37332.1; JOINED.
DR   EMBL; K03033; AAA37332.1; JOINED.
DR   EMBL; K03034; AAA37332.1; JOINED.
DR   EMBL; K03035; AAA37332.1; JOINED.
DR   PIR; A23982; A23982.
DR   MGD; MGI:88467; Cola1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD002078; Fib_collagen_C; 1.
```

```
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL       1     22
FT   PROPEP      23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN      152   1207       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP     1208   1453       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       29     87       VWFC.
FT   DOMAIN      152    167       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN      168   1181       TRIPLE-HELICAL REGION.
FT   DOMAIN     1182   1207       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1354   1354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        734    736       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE       1082   1084       CELL ATTACHMENT SITE (POTENTIAL).
FT   CONFLICT   1450   1450       A -> V (IN REF. 5).
SQ   SEQUENCE   1453 AA;  137944 MW;  3B802E535DF81808 CRC64;


  Query Match             96.4%;  Score 321;  DB 1;  Length 1453;
  Best Local Similarity   96.6%;  Pred. No. 1.3e-18;
  Matches   57;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||| |
Db     520 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGTA 578


RESULT   5
CA11_RAT
ID   CA11_RAT       STANDARD;      PRT;   671 AA.
AC   P02454; P02455;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN (FRAGMENTS).
GN   COL1A1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-19.
RX   MEDLINE=69155173; PubMed=5777344;
RA   Bornstein P.;
RT   "Comparative sequence studies of rat skin and tendon collagen. II.
RT   The absence of a short sequence at the amino terminus of the skin
RT   alpha-1 chain.";
RL   Biochemistry 8:63-71(1969).
RN   [2]
RP   SEQUENCE OF 5-19.
RX   MEDLINE=67162268; PubMed=5337886;
RA   Kang A.H., Bornstein P., Piez K.A.;
RT   "The amino acid sequence of peptides from the cross-linking region of
RT   rat skin collagen.";
RL   Biochemistry 6:788-795(1967).
RN   [3]
RP   SEQUENCE OF 20-55.
RX   MEDLINE=67165368; PubMed=4290711;
RA   Bornstein P.;
RT   "The incomplete hydroxylation of individual prolyl residues in
RT   collagen.";
RL   J. Biol. Chem. 242:2572-2574(1967).
RN   [4]
RP   SEQUENCE OF 56-102.
RX   MEDLINE=71263178; PubMed=4327399;
RA   Butler W.T., Ponds S.L.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. Amino acid sequence of alpha 1-CB4.";
RL   Biochemistry 10:2076-2081(1971).
RN   [5]
RP   SEQUENCE OF 103-139.
RX   MEDLINE=70085124; PubMed=5411206;
RA   Butler W.T.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. The covalent structure of alpha 1-CB5, the major
RT   hexose-containing cyanogen bromide peptide of alpha 1.";
RL   Biochemistry 9:44-50(1970).
RN   [6]
RP   SEQUENCE OF 140-238.
RX   MEDLINE=72136131; PubMed=4335087;
RA   Balian G., Click E.M., Bornstein P.;
RT   "Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of
RT   the hydroxylamine-produced fragment HA1.";
RL   Biochemistry 10:4470-4478(1971).
RN   [7]
RP   SEQUENCE OF 239-418.
RX   MEDLINE=73006942; PubMed=4342027;
RA   Balian G., Click E.M., Hermodson M.A., Bornstein P.;
RT   "Structure of rat skin collagen alpha 1-CBB. Amino acid sequence of
RT   the hydroxyl amine-produced fragment HA2.";
RL   Biochemistry 11:3798-3806(1972).
RN   [8]
RP   SEQUENCE OF 419-567.
RX   MEDLINE=74271984; PubMed=4366532;
RA   Butler W.T., Underwood S.P., Finch J.E. Jr.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. Amino acid sequence of alpha 1-CB3.";
RL   Biochemistry 13:2946-2953(1974).
RN   [9]
RP   SEQUENCE OF 568-651.
RX   MEDLINE=74011954; PubMed=4126850;
RA   Stoltz M., Timpl R., Furthmayr H., Kuehn K.;
RT   "Structural and immunogenic properties of a major antigenic
RT   determinant in neutral salt-extracted rat-skin collagen.";
RL   Eur. J. Biochem. 37:287-294(1973).
RN   [10]
RP   SEQUENCE OF 651-671.
RX   MEDLINE=73049495; PubMed=4636751;
RA   Stoltz M., Timpl R., Kuehn K.;
RT   "Non-helical regions in rat collagen alpha 1-chain.";
RL   FEBS Lett. 26:61-65(1972).
RN   [11]
RP   SEQUENCE OF 529-567 FROM N.A.
RX   MEDLINE=85122694; PubMed=6395893;
RA   Genovese C., Rowe D., Kream B.;
RT   "Construction of DNA sequences complementary to rat alpha 1 and alpha
RT   2 collagen mRNA and their use in studying the regulation of type I
RT   collagen synthesis by 1,25-dihydroxyvitamin D.";
RL   Biochemistry 23:6210-6216(1984).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC       O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11432; AAA40832.1; ALT_SEQ.
DR   PIR; A02854; CGRT1S.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 10.
DR   PROSITE; PS01208; VWFC; PARTIAL.
```

```
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT   MOD_RES       9      9       CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                                INVOLVED IN CROSS-LINKING.
FT   MOD_RES      28     28       HYDROXYLATION (PROBABLE).
FT   MOD_RES      31     31       HYDROXYLATION (PROBABLE).
FT   MOD_RES      34     34       HYDROXYLATION (PROBABLE).
FT   MOD_RES      43     43       HYDROXYLATION (PROBABLE).
FT   MOD_RES      46     46       HYDROXYLATION (PROBABLE).
FT   MOD_RES      49     49       HYDROXYLATION (PROBABLE).
FT   MOD_RES     103    103       HYDROXYLATION.
FT   CARBOHYD    103    103       O-LINKED (GAL. . .).
FT   MOD_RES     424    424       HYDROXYLATION (PROBABLE).
FT   MOD_RES     547    547       HYDROXYLATION (PROBABLE).
FT   NON_CONS    567    568
FT   DOMAIN      641    651       MAJOR ANTIGENIC DETERMINANT (OF NEUTRAL
FT                                SALT-EXTRACTED RAT SKIN COLLAGEN).
SQ   SEQUENCE   671 AA;  60615 MW;  9DC3114204AC4918 CRC64;


  Query Match             95.2%;  Score 317;  DB 1;  Length 671;
  Best Local Similarity   93.2%;  Pred. No. 1.4e-18;
  Matches   55;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
           ||||||||||||||||||||||||||||::||||| |||||||||||||||||| |
Db     369 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGZBGRPGPAGPPGARGQAGVMGFPGPKGTA 427


RESULT   6
CA12_BOVIN
ID   CA12_BOVIN     STANDARD;      PRT;   747 AA.
AC   P02459; Q28070; Q9XT24;
DT   21-JUL-1986 (Rel. 01, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(II) CHAIN PRECURSOR (FRAGMENTS).
GN   COL2A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-15.
RC   TISSUE=Cartilage;
RX   MEDLINE=73258693; PubMed=4732855;
RA   Miller E.J., Lunde L.G.;
RT   "Isolation and characterization of the cyanogen bromide peptides from
RT   the alpha 1(II) chain of bovine and human cartilage collagen.";
RL   Biochemistry 12:3153-3159(1973).
RN   [2]
RP   SEQUENCE OF 16-177.
RC   TISSUE=Cartilage;
RX   MEDLINE=76253504; PubMed=782511;
RA   Butler W.T., Miller E.J., Finch J.E. Jr.;
RT   "The covalent structure of cartilage collagen. Amino acid sequence of
RT   the NH2-terminal helical portion of the alpha 1 (II) chain.";
RL   Biochemistry 15:3000-3006(1976).
RN   [3]
RP   SEQUENCE OF 139-198.
RC   TISSUE=Cartilage;
RX   MEDLINE=77093864; PubMed=833147;
RA   Butler W.T., Finch J.E. Jr., Miller E.J.;
RT   "The covalent structure of cartilage collagen. Evidence for sequence
RT   heterogeneity of bovine alpha1(II) chains.";
RL   J. Biol. Chem. 252:639-643(1977).
RN   [4]
RP   SEQUENCE OF 139-417.
RC   TISSUE=Cartilage;
RX   MEDLINE=89231683; PubMed=2714276;
RA   Seyer J.M., Hasty K.A., Kang A.H.;
RT   "Covalent structure of collagen. Amino acid sequence of an
RT   arthritogenic cyanogen bromide peptide from type II collagen of
RT   bovine cartilage.";
RL   Eur. J. Biochem. 181:159-173(1989).
RN   [5]
RP   SEQUENCE OF 418-492.
RX   MEDLINE=74163168; PubMed=4857180;
RA   Butler W.T., Miller E.J., Finch J.E. Jr., Inagami T.;
RT   "Homologous regions of collagen alpha1(I) and alpha1(II) chains:
RT   apparent clustering of variable and invariant amino acid residues.";
RL   Biochem. Biophys. Res. Commun. 57:190-195(1974).
RN   [6]
RP   SEQUENCE OF 180-272 FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=94194070; PubMed=7511638;
RA   Brand D.D., Myers L.K., Terato K., Whittington K.B., Stuart J.M.,
RA   Rosloniec E.F.;
RT   "Characterization of the T cell determinants in the induction of
RT   autoimmune arthritis by bovine alpha 1(II)-CB11 in H-2q mice.";
RL   J. Immunol. 152:3088-3097(1994).
RN   [7]
RP   SEQUENCE OF 417-566 FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=99410731; PubMed=10479530;
RA   Tang B., Chiang T.M., Brand D.D., Gumanovskaya M.L., Stuart J.M.,
RA   Kang A.H., Myers L.K.;
RT   "Molecular definition and characterization of recombinant bovine CB8
RT   and CB10: immunogenicity and arthritogenicity.";
RL   Clin. Immunol. 92:256-264(1999).
RN   [8]
RP   SEQUENCE OF 567-747 FROM N.A.
RX   MEDLINE=85215651; PubMed=2582365;
RA   Sangiorgi F.O., Benson-Chanda V., de Wet W.J., Sobel M.E.,
RA   Ramirez F.;
RT   "Analysis of cDNA and genomic clones coding for the pro alpha 1 chain
RT   of calf type II collagen.";
RL   Nucleic Acids Res. 13:2815-2826(1985).
CC   -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC       O-LINKED GLYCANS CONSIST OF GLC-GAL DISACCHARIDES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L28918; AAA30436.1; -.
DR   EMBL; AF138957; AAD42347.1; -.
DR   EMBL; X02420; CAA26269.1; -.
DR   PIR; A02859; CGBO6C.
DR   PIR; A05039; A05039.
DR   PIR; S03940; S03940.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Cartilage; Collagen.
FT   CHAIN         1   >566       COLLAGEN ALPHA 1(II) CHAIN.
FT   NON_CONS    566    567
FT   PROPEP     <567    747       CARBOXYL-TERMINAL PROPEPTIDE.
FT   MOD_RES      9      9       HYDROXYLATION (INVOLVED IN CROSS-
FT                                LINKING).
FT   MOD_RES    102    102       HYDROXYLATION.
```

```
FT   CARBOHYD    102    102       O-LINKED (GAL. . .).
FT   MOD_RES     114    114       HYDROXYLATION.
FT   CARBOHYD    114    114       O-LINKED (GAL. . .).
FT   MOD_RES     123    123       HYDROXYLATION.
FT   CARBOHYD    123    123       O-LINKED (GAL. . .).
FT   MOD_RES     189    189       HYDROXYLATION.
FT   MOD_RES     423    423       HYDROXYLATION.
FT   CARBOHYD    423    423       O-LINKED (GAL).
FT   MOD_RES     435    435       HYDROXYLATION.
FT   CARBOHYD    435    435       O-LINKED (GAL).
FT   VARIANT     143    143       L -> A (IN MINOR COMPONENT).
FT   VARIANT     164    164       Q -> L (IN MINOR COMPONENT).
FT   CONFLICT    179    179       G -> Z (IN REF. 3).
FT   CONFLICT    185    186       AP -> PA (IN REF. 3).
FT   CONFLICT    191    192       EA -> AS (IN REF. 3).
FT   CONFLICT    195    195       T -> Q (IN REF. 4).
FT   CONFLICT    215    215       S -> A (IN REF. 4).
FT   CONFLICT    227    227       T -> A (IN REF. 4).
FT   CONFLICT    251    251       P -> A (IN REF. 4).
FT   CONFLICT    258    258       Q -> T (IN REF. 4).
FT   CONFLICT    261    261       T -> S (IN REF. 4).
FT   CONFLICT    492    492       G -> P (IN REF. 5).
SQ   SEQUENCE   747 AA;  71329 MW;  D0FC1D7CD1CAF77C CRC64;


  Query Match             79.3%;  Score 264;  DB 1;  Length 747;
  Best Local Similarity   81.0%;  Pred. No. 2.2e-14;
  Matches   47;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          |||||||||||| ||  || || || |  |:|||||||||| ||||| ||||||||||
Db    368 EAGLPGAKGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 425


RESULT   7
CA12_HUMAN
ID   CA12_HUMAN     STANDARD;      PRT;  1418 AA.
AC   P02458;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90067946; PubMed=2587267;
RA   Su M.W., Lee B., Ramirez F., Machado M., Horton W.;
RT   "Nucleotide sequence of the full length cDNA encoding for human type
RT   II procollagen.";
RL   Nucleic Acids Res. 17:9473-9473(1989).
RN   [2]
RP   SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE=87031574; PubMed=3021582;
RA   Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
RT   "Promoter region of the human pro-alpha 1(II)-collagen gene.";
RL   Gene 44:11-16(1986).
RN   [3]
RP   SEQUENCE OF 432-1145 FROM N.A.
RA   Ramirez F.;
RL   Submitted (DEC-1988) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 963-1418 FROM N.A.
RX   MEDLINE=85190534; PubMed=3857598;
RA   Cheah K.S.E., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
RT   "Identification and characterization of the human type II collagen
RT   gene (COL2A1).";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN   [5]
RP   SEQUENCE OF 1120-1398 FROM N.A.
RX   MEDLINE=85306861; PubMed=3840017;
RA   Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,
RA   Vuorio E.;
RT   "Construction and identification of a cDNA clone for human type II
RT   procollagen mRNA.";
RL   Biochem. J. 229:183-188(1985).
RN   [6]
RP   SEQUENCE OF 1106-1418 FROM N.A.
RX   MEDLINE=88067771; PubMed=2825137;
RA   Elima K., Vuorio T., Vuorio E.;
RT   "Determination of the single polyadenylation site of the human pro
RT   alpha 1(II) collagen gene.";
RL   Nucleic Acids Res. 15:9499-9504(1987).
RN   [7]
RP   SEQUENCE OF 1227-1289 FROM N.A.
RX   MEDLINE=86104139; PubMed=3002437;
RA   Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
RT   "Isolation and partial characterization of genomic clones coding for
RT   a human pro-alpha 1 (II) collagen chain and demonstration of
RT   restriction fragment length polymorphism at the 3' end of the gene.";
RL   Biochemistry 24:6343-6348(1985).
RN   [8]
RP   SEQUENCE OF 1176-1226 FROM N.A.
RX   MEDLINE=84118798; PubMed=6320112;
RA   Strom C.M., Upholt W.B.;
RT   "Isolation and characterization of genomic clones corresponding to
RT   the human type II procollagen gene.";
RL   Nucleic Acids Res. 12:1025-1038(1984).
RN   [9]
RP   SEQUENCE OF 35-167 FROM N.A.
RX   MEDLINE=89233138; PubMed=2714801;
RA   Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
RT   "Organization of the exons coding for pro alpha 1(II) collagen N-
RT   propeptide confirms a distinct evolutionary history of this domain of
RT   the fibrillar collagen genes.";
RL   Genomics 4:438-441(1989).
RN   [10]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=91184577; PubMed=2010058;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in collagen genes: causes of rare and some common diseases
RT   in humans.";
RL   FASEB J. 5:2052-2060(1991).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [12]
RP   VARIANT SER-1074.
RX   MEDLINE=90036909; PubMed=2572591;
RA   Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
RA   Hollister D.W.;
RT   "Glycine to serine substitution in the triple helical domain of pro-
RT   alpha 1 (II) collagen results in a lethal perinatal form of short-
RT   limbed dwarfism.";
RL   J. Biol. Chem. 264:18265-18267(1989).
RN   [13]
RP   VARIANT SEDC GLY-1095--TYR-1330 DEL.
RX   MEDLINE=89266907; PubMed=2543071;
RA   Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
RT   "Identification of the molecular defect in a family with
RT   spondyloepiphyseal dysplasia.";
RL   Science 244:978-980(1989).
RN   [14]
RP   VARIANT OSTEOARTHRITIS CYS-650.
RX   MEDLINE=90370826; PubMed=1975693;
RA   Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
RT   "Single base mutation in the type II procollagen gene (COL2A1) as a
RT   cause of primary osteoarthritis associated with a mild
```

```
RT   chondrodysplasia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
RN   [15]
RP   VARIANT OI-IV VAL-717.
RX   MEDLINE=91291136; PubMed=2064612;
RA   Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RT   "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT   valine substitution in osteogenesis imperfecta type IV. Detection of
RT   the mutation and prenatal diagnosis by a chemical cleavage method.";
RL   Biochem. J. 276:765-770(1991).
RN   [16]
RP   VARIANT OSTEOARTHRITIS CYS-650.
RX   MEDLINE=91086471; PubMed=1985108;
RA   Eyre D.R., Weis M.A., Moskowitz R.W.;
RT   "Cartilage expression of a type II collagen mutation in an inherited
RT   form of osteoarthritis associated with a mild chondrodysplasia.";
RL   J. Clin. Invest. 87:357-361(1991).
RN   [17]
RP   VARIANT HYPOCHONDROGENESIS GLU-984.
RX   MEDLINE=93054548; PubMed=1429602;
RA   Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
RA   Cohn D.H., Eyre D.R.;
RT   "An amino acid substitution (Gly853-->Glu) in the collagen alpha
RT   1(II) chain produces hypochondrogenesis.";
RL   J. Biol. Chem. 267:22522-22526(1992).
RN   [18]
RP   VARIANT HYPOCHONDROGENESIS SER-705.
RX   MEDLINE=92262484; PubMed=1374906;
RA   Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
RA   Ramirez F., Vitale E., Lee B.;
RT   "Characterization of a type II collagen gene (COL2A1) mutation
RT   identified in cultured chondrocytes from human hypochondrogenesis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
RN   [19]
RP   VARIANT WS-II ASP-198.
RX   MEDLINE=93304428; PubMed=8317498;
RA   Koerkkoe J., Ritvaniemi P., Haataja L., Kaeaeriaeinen H.,
RA   Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
RT   "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT   for glycine alpha 1-67 and that causes cataracts and retinal
RT   detachment: evidence for molecular heterogeneity in the Wagner
RT   syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL   Am. J. Hum. Genet. 53:55-61(1993).
RN   [20]
RP   VARIANT SEMD CYS-840.
RA   Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
RA   Eyre D.R.;
RT   "A dominant mutation in the type II collagen gene (COL2A1) produces
RT   spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
RL   Am. J. Hum. Genet. 53:A209-A209(1993).
RN   [21]
RP   VARIANT OSTEOARTHRITIS CYS-650.
RX   MEDLINE=93282819; PubMed=8507190;
RA   Holderbaum D., Malemud C.J., Moskowitz R.W., Haqqi T.M.;
RT   "Human cartilage from late stage familial osteoarthritis transcribes
RT   type II collagen mRNA encoding a cysteine in position 519.";
RL   Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
RN   [22]
RP   VARIANT SEMD ARG-285.
RX   MEDLINE=93252400; PubMed=8486375;
RA   Vikkula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
RA   Peltonen L.;
RT   "A mutation in the amino-terminal end of the triple helix of type II
RT   collagen causing severe osteochondrodysplasia.";
RL   Genomics 16:282-285(1993).
RN   [23]
RP   VARIANT SEDC CYS-206.
RX   MEDLINE=94063862; PubMed=8244341;
RA   Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
RA   Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
RT   "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
RT   family with an Arg75-->Cys mutation in the procollagen type II gene
RT   (COL2A1).";
RL   Hum. Genet. 92:499-505(1993).
RN   [24]
RP   VARIANT SEDC CYS-920.
RX   MEDLINE=93315508; PubMed=8325895;
RA   Chan D., Taylor T.K.F., Cole W.G.;
RT   "Characterization of an arginine 789 to cysteine substitution in
RT   alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
RT   dysplasia.";
RL   J. Biol. Chem. 268:15238-15245(1993).
RN   [25]
RP   VARIANT SEDC SER-1128.
RX   MEDLINE=93140139; PubMed=8423604;
RA   Cole W.G., Hall R.K., Rogers J.G.;
RT   "The clinical features of spondyloepiphyseal dysplasia congenita
RT   resulting from the substitution of glycine 997 by serine in the alpha
RT   1(II) chain of type II collagen.";
RL   J. Med. Genet. 30:27-35(1993).

  Query Match             76.9%;  Score 256;  DB 1;  Length 1418;
  Best Local Similarity   77.6%;  Pred. No. 1.6e-13;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | |||||:|||| ||  || || || |  |:|||||||||| ||||| ||||||||||
Db     484 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 541


RESULT   8
CA12_MOUSE
ID   CA12_MOUSE     STANDARD;      PRT;  1459 AA.
AC   P28481;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].
GN   COL2A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91358489; PubMed=1885613;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. Biol. Chem. 266:16862-16869(1991).
RN   [2]
RP   SEQUENCE OF 1455-1459 FROM N.A.
RX   MEDLINE=91274355; PubMed=2054384;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Specific hybridization probes for mouse type I, II, III and IX
RT   collagen mRNAs.";
RL   Biochim. Biophys. Acta 1089:241-243(1991).
CC   -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M65161; AAA68100.1; -.
DR   EMBL; X57982; CAA41047.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
```

```
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   PROPEP       26    153       AMINO-TERMINAL PROPEPTIDE
FT                                (BY SIMILARITY).
FT   CHAIN       154   1213       COLLAGEN ALPHA 1(II) CHAIN.
FT   PROPEP     1214   1459       CARBOXYL-TERMINAL PROPEPTIDE
FT                                (CHONDROCALCIN).
FT   DOMAIN       32     89       VWFC.
FT   DOMAIN      173   1186       TRIPLE-HELICAL REGION.
FT   DOMAIN     1187   1213       NONHELICAL REGION (C-TERMINAL).
FT   VARSPLIC     29     29       Q -> R (IN SHORT ISOFORM).
FT   VARSPLIC     30     98       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   1459 AA;  139154 MW;  F6C84FA7C532E7F2 CRC64;


  Query Match             76.9%;  Score 256;  DB 1;  Length 1459;
  Best Local Similarity   77.6%;  Pred. No. 1.6e-13;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | |||||:|||| ||  || || || |  |:|||||||||| ||||| ||||||||||
Db     525 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 582


RESULT  9
CA13_HUMAN
ID   CA13_HUMAN     STANDARD;      PRT;  1466 AA.
AC   P02461; Q15112;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(III) CHAIN PRECURSOR.
GN   COL3A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89350838; PubMed=2764886;
RA   Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA   Prockop D.J.;
RT   "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT   chain of human type III procollagen. Differences in protein structure
RT   from type I procollagen and conservation of codon preferences.";
RL   Biochem. J. 260:509-516(1989).
RN   [2]
RP   SEQUENCE OF 149-1225 FROM N.A.
RX   MEDLINE=89386015; PubMed=2780304;
RA   Janeczko R.A., Ramirez F.;
RT   "Nucleotide and amino acid sequences of the entire human alpha 1
RT   (III) collagen.";
RL   Nucleic Acids Res. 17:6742-6742(1989).
RN   [3]
RP   SEQUENCE OF 168-398.
RX   MEDLINE=77134724; PubMed=557335;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of cyanogen
RT   bromide peptides from the amino-terminal segment of type III collagen
RT   of human liver.";
RL   Biochemistry 16:1158-1164(1977).
RN   [4]
RP   REVISIONS.
RA   Seyer J.M.;
RL   Submitted (DEC-1977) to the PIR data bank.
RN   [5]
RP   SEQUENCE OF 399-727.
RX   MEDLINE=79000343; PubMed=687591;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of five
RT   consecutive CNBr peptides from type III collagen of human liver.";
RL   Biochemistry 17:3404-3411(1978).
RN   [6]
RP   SEQUENCE OF 728-964.
RX   MEDLINE=80198282; PubMed=6246925;
RA   Seyer J.M., Mainardi C., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha 1
RT   (III)-CB5 from type III collagen of human liver.";
RL   Biochemistry 19:1583-1589(1980).
RN   [7]
RP   SEQUENCE OF 950-1466 FROM N.A.
RX   MEDLINE=88189827; PubMed=3357782;
RA   Mankoo B.S., Dalgleish R.;
RT   "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL   Nucleic Acids Res. 16:2337-2337(1988).
RN   [8]
RP   REVISION TO 1184.
RX   MEDLINE=89098346; PubMed=3211760;
RA   Molyneux K., Dalgleish R.;
RT   "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL   Nucleic Acids Res. 16:11833-11833(1988).
RN   [9]
RP   SEQUENCE OF 1065-1466 FROM N.A.
RX   MEDLINE=85087944; PubMed=6096827;
RA   Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA   Rosenbloom J., Myers J.C.;
RT   "Molecular cloning and carboxyl-propeptide analysis of human type III
RT   procollagen.";
RL   Nucleic Acids Res. 12:9383-9394(1984).
RN   [10]
RP   SEQUENCE OF 965-1200.
RX   MEDLINE=81208139; PubMed=7016180;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha
RT   1(III)-CB9 from type III collagen of human liver.";
RL   Biochemistry 20:2621-2627(1981).
RN   [11]
RP   SEQUENCE OF 1176-1466 FROM N.A.
RX   MEDLINE=85157600; PubMed=2579949;
RA   Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT   "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT   (III) collagen. Partial characterization of the 3' end region of the
RT   gene.";
RL   J. Biol. Chem. 260:4357-4363(1985).
RN   [12]
RP   SEQUENCE OF 1161-1200 FROM N.A.
RX   MEDLINE=86187804; PubMed=3754462;
RA   Miskulin M., Dalgleish R., Kluve-Beckerman B., Rennard S.I.,
RA   Tolstoshev P., Brantly M., Crystal R.G.;
RT   "Human type III collagen gene expression is coordinately modulated
RT   with the type I collagen genes during fibroblast growth.";
RL   Biochemistry 25:1408-1413(1986).
RN   [13]
RP   SEQUENCE OF 1-170 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=88303360; PubMed=3405773;
RA   Toman D., Ricca G., de Crombrugghe B.;
RT   "Nucleotide sequence of a cDNA coding for the amino-terminal region
RT   of human prepro alpha 1(III) collagen.";
RL   Nucleic Acids Res. 16:7201-7201(1988).
RN   [14]
RP   SEQUENCE OF 1-176 FROM N.A.
RX   MEDLINE=89378752; PubMed=2777083;
RA   Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT   "Cloning and analysis of the 5' portion of the human type-III
RT   procollagen gene (COL3A1).";
```

```
RL   Gene 78:255-265(1989).
RN   [15]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [16]
RP   VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
RX   MEDLINE=93293988; PubMed=8514866;
RA   Tromp G., Wu Y., Prockop D.J., Madhatheri S.L., Kleinert C.,
RA   Earley J.J., Zhuang J., Noerrgaard O., Darling R.C., Abbott W.M.,
RA   Cole C.W., Jaakkola P., Ryynaenen M., Pearce W.H., Yao J.S.T.,
RA   Majamaa K., Smullens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
RA   Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
RT   "Sequencing of cDNA from 50 unrelated patients reveals that mutations
RT   in the triple-helical domain of type III procollagen are an
RT   infrequent cause of aortic aneurysms.";
RL   J. Clin. Invest. 91:2539-2545(1993).
RN   [17]
RP   VARIANT THR-698.
RX   MEDLINE=91045136; PubMed=2235526;
RA   Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
RA   Wu Y., Ganguly A., Prockop D.J.;
RT   "G to A polymorphism in exon 31 of the COL3A1 gene.";
RL   Nucleic Acids Res. 18:6180-6180(1990).
RN   [18]
RP   VARIANT AORTIC ANEURYSM ARG-786.
RX   MEDLINE=91056145; PubMed=2243125;
RA   Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
RT   "A mutation in the gene for type III procollagen (COL3A1) in a family
RT   with aortic aneurysms.";
RL   J. Clin. Invest. 86:1465-1473(1990).
RN   [19]
RP   VARIANT EDS-IV ARG-828.
RX   MEDLINE=94016385; PubMed=8411057;
RA   Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
RT   "The substitution of glycine 661 by arginine in type III collagen
RT   produces mutant molecules with different thermal stabilities and
RT   causes Ehlers-Danlos syndrome type IV.";
RL   J. Med. Genet. 30:690-693(1993).
RN   [20]
RP   VARIANT EDS-IV SER-957.
RX   MEDLINE=89109135; PubMed=2492273;
RA   Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
RT   "A single base mutation that substitutes serine for glycine 790 of
RT   the alpha 1 (III) chain of type III procollagen exposes an arginine
RT   and causes Ehlers-Danlos syndrome IV.";
RL   J. Biol. Chem. 264:1349-1352(1989).
RN   [21]
RP   VARIANT EDS-IV VAL-960.
RX   MEDLINE=95268429; PubMed=7749417;
RA   Tromp G., de Paepe A., Nuytinck L., Madhatheri S.L., Kuivaniemi H.;
RT   "Substitution of valine for glycine 793 in type III procollagen in
RT   Ehlers-Danlos syndrome type IV.";
RL   Hum. Mutat. 5:179-181(1995).
RN   [22]
RP   VARIANT EDS-IV GLU-1014.
RX   MEDLINE=92316511; PubMed=1352273;
RA   Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
RA   Pope F.M.;
RT   "A single base mutation in the gene for type III collagen (COL3A1)
RT   converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT   syndrome type IV. An unaffected family member is mosaic for the
RT   mutation.";
RL   Hum. Genet. 89:414-418(1992).
RN   [23]
RP   VARIANT EDS-IV ASP-1050.
RX   MEDLINE=90037070; PubMed=2808425;
RA   Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
RT   "Single base mutation in the type III procollagen gene that converts
RT   the codon for glycine 883 to aspartate in a mild variant of
RT   Ehlers-Danlos syndrome IV.";
RL   J. Biol. Chem. 264:19313-19317(1989).
RN   [24]
RP   VARIANT EDS-IV VAL-1077.
RX   MEDLINE=91374480; PubMed=1895316;
RA   Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,
RA   Pope F.M.;
RT   "Characterisation of a glycine to valine substitution at amino acid
RT   position 910 of the triple helical region of type III collagen in a
RT   patient with Ehlers-Danlos syndrome type IV.";
RL   J. Med. Genet. 28:458-463(1991).
RN   [25]
RP   VARIANT EDS-IV GLU-1173.
RX   MEDLINE=93022543; PubMed=1357232;
RA   Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;

  Query Match             70.3%;  Score 234;  DB 1;  Length 1466;
  Best Local Similarity   72.7%;  Pred. No. 8.8e-12;
  Matches   40;  Conservative    3;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           | || :|: |||| || ||| |||| |: |||||||| | ||| ||||||||||
Db     531 GGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSGPRGQPGVMGFPGPKG 585


RESULT  10
CA13_MOUSE
ID   CA13_MOUSE     STANDARD;      PRT;  1464 AA.
AC   P08121; Q61429; Q9CRN7;
DT   01-AUG-1988 (Rel. 08, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(III) CHAIN PRECURSOR.
GN   COL3A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL X DBA; TISSUE=Embryo;
RX   MEDLINE=95011609; PubMed=7926795;
RA   Toman D., de Crombrugghe B.;
RT   "The mouse type-III procollagen-encoding gene: genomic cloning and
RT   complete DNA sequence.";
RL   Gene 147:161-168(1994).
RN   [2]
RP   SEQUENCE OF 1-488 FROM N.A.
RX   MEDLINE=88167858; PubMed=3443309;
RA   Wood L., Theriault N., Vogeli G.;
RT   "Complete nucleotide sequence of the N-terminal domains of the murine
RT   alpha-1 type-III collagen chain.";
RL   Gene 61:225-230(1987).
RN   [3]
RP   SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE=85131189; PubMed=3972847;
RA   Liau G., Mudryj M., de Crombrugghe B.;
RT   "Identification of the promoter and first exon of the mouse alpha 1
RT   (III) collagen gene.";
RL   J. Biol. Chem. 260:3773-3777(1985).
RN   [4]
RP   SEQUENCE OF 810-1464 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
```

```
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [5]
RP   SEQUENCE OF 1442-1464 FROM N.A.
RC   STRAIN=C57BL;
RX   MEDLINE=91274355; PubMed=2054384;
RA   Metsaeranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Specific hybridization probes for mouse type I, II, III and IX
RT   collagen mRNAs.";
RL   Biochim. Biophys. Acta 1089:241-243(1991).
CC   -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC       ALONG WITH TYPE I COLLAGEN.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC       LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC       ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC       O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52046; CAA36279.1; -.
DR   EMBL; M18933; AAA37338.1; -.
DR   EMBL; K03037; -; NOT_ANNOTATED_CDS.
DR   EMBL; AK019448; BAB31724.1; -.
DR   EMBL; X57983; CAA41048.1; -.
DR   PIR; A22287; A22287.
DR   PIR; A27353; A27353.
DR   PIR; S16373; S16373.
DR   MGD; MGI:88453; Col3a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL       1     23       BY SIMILARITY.
FT   PROPEP      24    154       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN      155   1203       COLLAGEN ALPHA 1(III) CHAIN.
FT   PROPEP    1204   1464       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN      31     90       VWFC.
FT   DOMAIN     155    169       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN     170   1195       TRIPLE-HELICAL REGION.
FT   DOMAIN    1196   1464       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD   262    262       O-LINKED (GAL. . .) (BY SIMILARITY).
FT   MOD_RES    262    262       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    283    283       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    859    859       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    976    976       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES   1093   1093       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES   1105   1105       HYDROXYLATION (BY SIMILARITY).
FT   DISULFID  1195   1195       INTERCHAIN (BY SIMILARITY).
FT   DISULFID  1196   1196       INTERCHAIN (BY SIMILARITY).
SQ   SEQUENCE   1464 AA;  138944 MW;  2104EC27A886090B CRC64;


  Query Match             70.0%;  Score 233;  DB 1;  Length 1464;
  Best Local Similarity   72.7%;  Pred. No. 1.1e-11;
  Matches   40;  Conservative    3;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           | || :|: |||| || ||| |||| |: ||||||| | ||| ||||||||||
Db     530 GGPGIRGMPGSPGGPGNDGKPGPPGSQGESGRPGPPGPSGPRGQPGVMGFPGPKG 584


RESULT  11
CA13_CHICK
ID   CA13_CHICK     STANDARD;      PRT;  1262 AA.
AC   P12105; P79758; P79759; Q90794; Q92029;
DT   01-OCT-1989 (Rel. 12, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(III) CHAIN PRECURSOR (FRAGMENTS).
GN   COL3A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-886 FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=94266842; PubMed=8206952;
RA   Nah H.-D., Niu Z., Adams S.L.;
RT   "An alternative transcript of the chick type III collagen gene that
RT   does not encode type III collagen.";
RL   J. Biol. Chem. 269:16443-16448(1994).
RN   [2]
RP   SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.
RX   MEDLINE=84270696; PubMed=6547770;
RA   Yamada Y., Liau G., Mudryj M., Obici S., de Crombrugghe B.;
RT   "Conservation of the sizes for one but not another class of exons in
RT   two chick collagen genes.";
RL   Nature 310:333-337(1984).
RN   [3]
RP   SEQUENCE OF 977-1262 FROM N.A.
RX   MEDLINE=83220816; PubMed=6856474;
RA   Yamada Y., Kuhn K., de Crombrugghe B.;
RT   "A conserved nucleotide sequence, coding for a segment of the C-
RT   propeptide, is found at the same location in different collagen
RT   genes.";
RL   Nucleic Acids Res. 11:2733-2744(1983).
CC   -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC       ALONG WITH TYPE I COLLAGEN.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC       LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC       ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U07973; AAA83407.1; -.
DR   EMBL; X00822; CAB52686.1; -.
DR   EMBL; X00823; CAB52686.1; JOINED.
DR   EMBL; X00824; CAA25396.1; ALT_SEQ.
DR   EMBL; X00823; CAA25396.1; JOINED.
```

```
DR   EMBL; X00822; CAA25396.1; JOINED.
DR   EMBL; X00826; CAA25397.1; ALT_SEQ.
DR   EMBL; X00825; CAA25397.1; JOINED.
DR   EMBL; X00827; CAA25398.1; -.
DR   EMBL; X00828; CAA25399.1; -.
DR   EMBL; X00830; CAA25401.1; -.
DR   EMBL; X00831; CAA25402.1; -.
DR   EMBL; K02302; AAD15299.1; -.
DR   EMBL; K02301; AAD15298.1; -.
DR   EMBL; M36662; AAA18519.1; ALT_SEQ.
DR   PIR; A05269; A05269.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 11.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24    144       AMINO-TERMINAL PROPEPTIDE (BY
FT                                SIMILARITY).
FT   CHAIN       145   1003       COLLAGEN ALPHA 1(III) CHAIN.
FT   PROPEP     1004   1262       CARBOXYL-TERMINAL PROPEPTIDE (BY
FT                                SIMILARITY).
FT   DOMAIN        29     88       VWFC.
FT   DOMAIN       145    164       NONHELICAL REGION (N-TERMINAL) (BY
FT                                SIMILARITY).
FT   DOMAIN       165    994       TRIPLE-HELICAL REGION (BY SIMILARITY).
FT   DOMAIN       995   1003       NONHELICAL REGION (C-TERMINAL) (BY
FT                                SIMILARITY).
FT   NON_CONS     886    887
FT   NON_CONS     922    923
FT   DISULFID     994    994       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     995    995       INTERCHAIN (BY SIMILARITY).
FT   MOD_RES      262    262       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES      283    283       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES      859    859       HYDROXYLATION (BY SIMILARITY).
FT   CARBOHYD    1163   1163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      96     96       E -> K (IN REF. 2).
FT   CONFLICT    1132   1132       F -> S (IN REF. 3).
SQ   SEQUENCE   1262 AA;  121249 MW;  96ABE7B2E9DEB43D CRC64;


  Query Match             69.7%;  Score 232;  DB 1;  Length 1262;
  Best Local Similarity   74.5%;  Pred. No. 1.1e-11;
  Matches   41;  Conservative    2;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           |||| :|| | ||||| ||| ||||  |: || ||||| | ||| ||||||||||
Db     530 GLPGMRGLPGIPGSPGSDGKPGPPGNQGEPGRSGPPGPAGPRGQPGVMGFPGPKG 584


RESULT  12
CA13_BOVIN
ID   CA13_BOVIN     STANDARD;      PRT;  1049 AA.
AC   P04258;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(III) CHAIN.
GN   COL3A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-242.
RX   MEDLINE=80026026; PubMed=488906;
RA   Fietzek P.P., Allmann H., Rauterberg J., Henkel W., Wachter E.,
RA   Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. I. The amino
RT   acid sequence of the amino terminal region of the alpha 1(III) chain
RT   (positions 1-222).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
RN   [2]
RP   SEQUENCE OF 243-422.
RX   MEDLINE=80026027; PubMed=488907;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. II. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
RT   (positions 223-402).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
RN   [3]
RP   SEQUENCE OF 423-571.
RX   MEDLINE=80026028; PubMed=488908;
RA   Bentz H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. III. The
RT   amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT   (positions 403-551).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
RN   [4]
RP   SEQUENCE OF 572-808.
RX   MEDLINE=80026029; PubMed=488909;
RA   Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. IV. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT   (positions 552-788).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
RN   [5]
RP   SEQUENCE OF 809-947.
RX   MEDLINE=80026030; PubMed=488910;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. V. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT   (position 789-927).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
RN   [6]
RP   SEQUENCE OF 948-1049.
RX   MEDLINE=80026031; PubMed=488911;
RA   Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. VI. The amino
RT   acid sequence of the carboxyterminal cyanogen bromide peptide alpha
RT   1(III)CB9B (positions 928-1028).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
CC   -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC       ALONG WITH TYPE I COLLAGEN.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC       LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC       ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
DR   PIR; A02862; CGBO7S.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 17.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen.
FT   DOMAIN        1     14       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN       15   1040       TRIPLE-HELICAL REGION.
FT   DOMAIN     1041   1049       NONHELICAL REGION (C-TERMINAL).
FT   MOD_RES      95     95       HYDROXYLATION.
FT   MOD_RES     107    107       HYDROXYLATION.
FT   MOD_RES     119    119       HYDROXYLATION.
FT   MOD_RES     938    938       HYDROXYLATION.
FT   MOD_RES     950    950       HYDROXYLATION.
FT   CARBOHYD    107    107       O-LINKED (GAL. . .).
FT   CARBOHYD    950    950       O-LINKED (GAL. . .).
FT   DISULFID   1040   1040       INTERCHAIN.
FT   DISULFID   1041   1041       INTERCHAIN.
```

```
SQ   SEQUENCE   1049 AA;  93651 MW;  8EEC33D1C66EC9A3 CRC64;


  Query Match             68.5%;  Score 228;  DB 1;  Length 1049;
  Best Local Similarity   70.9%;  Pred. No. 2e-11;
  Matches   39;  Conservative    4;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 GLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           | || :|: |||| || :|| ||||  |: ||||||| || ||| ||||||||||
Db     375 GGPGLRGIPGSPGGPGSNGKPGPPGSQGETGRPGPPGSPGPRGQPGVMGFPGPKG 429


RESULT  13
CA25_HUMAN
ID   CA25_HUMAN     STANDARD;      PRT;  1496 AA.
AC   P05997;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 2(V) CHAIN PRECURSOR.
GN   COL5A2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-463 FROM N.A.
RX   MEDLINE=89123368; PubMed=2914927;
RA   Woodbury D., Benson-Chanda V., Ramirez F.;
RT   "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
RT   to the structural criteria of a fibrillar procollagen molecule.";
RL   J. Biol. Chem. 264:2735-2738(1989).
RN   [2]
RP   SEQUENCE OF 398-1496 FROM N.A.
RX   MEDLINE=87146331; PubMed=3029669;
RA   Weil D., Bernard M.P., Gargano S., Ramirez F.;
RT   "The pro alpha 2(V) collagen gene is evolutionarily related to the
RT   major fibrillar-forming collagens.";
RL   Nucleic Acids Res. 15:181-198(1987).
RN   [3]
RP   SEQUENCE OF 1227-1496 FROM N.A.
RX   MEDLINE=85289337; PubMed=2411731;
RA   Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
RT   "Complete primary structure of the human alpha 2 type V procollagen
RT   COOH-terminal propeptide.";
RL   J. Biol. Chem. 260:11216-11222(1985).
RN   [4]
RP   SEQUENCE OF 1449-1496 FROM N.A.
RX   MEDLINE=89138450; PubMed=3224983;
RA   Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
RA   Ramirez F.;
RT   "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
RT   located on the long arm of human chromosome 2.";
RL   Genomics 3:275-277(1988).
RN   [5]
RP   SEQUENCE OF 208-227.
RC   TISSUE=Placenta;
RX   MEDLINE=92239022; PubMed=1571108;
RA   Mann K.;
RT   "Isolation of the alpha 3-chain of human type V collagen and
RT   characterization by partial sequencing.";
RL   Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN   [6]
RP   SEQUENCE OF 288-297 AND 606-617.
RC   TISSUE=Bone;
RX   MEDLINE=94237164; PubMed=8181482;
RA   Moradi-Ameli M., Rousseau J.C., Kleman J.P., Champliaud M.F.,
RA   Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT   "Diversity in the processing events at the N-terminus of type-V
RT   collagen.";
RL   Eur. J. Biochem. 221:987-995(1994).
CC   -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE
CC       COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
CC       TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
CC   -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN
CC       MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
CC       ONE ALPHA 3(V) CHAINS IN PLACENTA.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04478; AAA51859.1; -.
DR   EMBL; X04758; CAA28454.1; -.
DR   EMBL; M11718; AAA52058.1; -.
DR   PIR; A25374; A25374.
DR   PIR; A25874; A25874.
DR   PIR; A30017; A30017.
DR   PIR; A31427; A31427.
DR   MIM; 120190; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     26
FT   CHAIN        27   1226       COLLAGEN ALPHA 2(V) CHAIN.
FT   PROPEP     1227   1496       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       39     97       VWFC.
FT   MOD_RES     290    290       HYDROXYLATION.
FT   MOD_RES     293    293       HYDROXYLATION.
FT   MOD_RES     296    296       HYDROXYLATION.
FT   MOD_RES     608    608       HYDROXYLATION.
FT   MOD_RES     614    614       HYDROXYLATION.
FT   CONFLICT    292    292       A -> P (IN REF. 6).
FT   CONFLICT   1418   1418       K -> T (IN REF. 3).
FT   CONFLICT   1438   1438       F -> S (IN REF. 3).
FT   CONFLICT   1460   1460       E -> Q (IN REF. 4).
FT   CONFLICT   1496   1496       V -> A (IN REF. 4).
SQ   SEQUENCE   1496 AA;  144720 MW;  82827C17A8644F5A CRC64;


  Query Match             67.3%;  Score 224;  DB 1;  Length 1496;
  Best Local Similarity   67.2%;  Pred. No. 5.5e-11;
  Matches   39;  Conservative    6;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | |||||:||||:||  ||:|| || |  |:||||||||  | :|| | || |||||:
Db     562 EPGLPGARGLTGNPGVQGPEGKLGPLGAPGEDGRPGPPGSIGIKGQPGTMGLPGPKGS 619


RESULT  14
CA21_ONCMY
ID   CA21_ONCMY     STANDARD;      PRT;  1356 AA.
AC   O93484;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 2(I) CHAIN PRECURSOR.
GN   COL1A2.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=11358497;
RA   Saito M., Takenouchi Y., Kunisaki N., Kimura S.;
RT   "Complete primary structure of rainbow trout type I collagen
RT   consisting of alpha1(I)alpha2(I)alpha3(I) heterotrimers.";
RL   Eur. J. Biochem. 268:2817-2827(2001).
RN   [2]
RP   SEQUENCE OF 417-1356 FROM N.A.
RC   TISSUE=Fibroblast;
RA   Saito M., Kunisaki N., Hirono I., Aoki T., Ishida M., Urano N.,
RA   Kimura S.;
RT   "Partial characterization of cDNA clones encoding the three distinct
RT   pro alpha chains of type I collagen from rainbow trout.";
RL   Fisheries Sci. 64:780-786(1998).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB052837; BAB55663.1; -.
DR   EMBL; AB008372; BAA33379.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 11.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   PROPEP       25      ?       AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
FT   CHAIN         ?   1096       COLLAGEN ALPHA 2(I) CHAIN.
FT   PROPEP     1097   1356       CARBOXYL-TERMINAL PROPEPTIDE
FT                                (BY SIMILARITY).
FT   CARBOHYD   1257   1257       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1356 AA;  126985 MW;  7BB2F1F80DB10C93 CRC64;


  Query Match             64.9%;  Score 216;  DB 1;  Length 1356;
  Best Local Similarity   68.4%;  Pred. No. 2.1e-10;
  Matches   39;  Conservative    4;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           |:|| ||:|| |: |  || || |||| || ||| ||||| | ||| | :|||||||
Db     437 ESGLTGARGLPGNSGQGGPPGKEGPPGAAGLDGRTGPPGPTGPRGQPGNIGFPGPKG 493


RESULT  15
CA21_MOUSE
ID   CA21_MOUSE     STANDARD;      PRT;  1372 AA.
AC   Q01149;
DT   01-APR-1993 (Rel. 25, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 2(I) CHAIN PRECURSOR.
GN   COL1A2 OR COLA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Calvaria;
RX   MEDLINE=92372043; PubMed=1505972;
RA   Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;
RT   "Sequence analysis of a full-length cDNA for the murine pro alpha
RT   2(I) collagen chain: comparison of the derived primary structure with
RT   human pro alpha 2(I) collagen.";
RL   Genomics 13:1345-1346(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast tumor;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-110 FROM N.A.
RC   TISSUE=Calvaria;
RX   MEDLINE=92084969; PubMed=1748823;
RA   Phillips C.L., Lever L.W., Pinnell S.R., Quarles L.D.,
RA   Wenstrup R.J.;
RT   "Construction of a full-length murine pro alpha 2(I) collagen cDNA by
RT   the polymerase chain reaction.";
RL   J. Invest. Dermatol. 97:980-984(1991).
RN   [4]
RP   SEQUENCE OF 1-23 FROM N.A.
RX   MEDLINE=87289650; PubMed=3039494;
RA   Rossi P., de Crombrugghe B.;
RT   "Identification of a cell-specific transcriptional enhancer in the
RT   first intron of the mouse alpha 2 (type I) collagen gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594(1987).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58251; CAA41205.1; -.
DR   EMBL; BC007158; AAH07158.1; -.
DR   EMBL; K01832; AAA37331.1; -.
DR   PIR; A43291; A43291.
DR   MGD; MGI:88468; Cola2.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     85       AMINO-TERMINAL PROPEPTIDE
FT                                (BY SIMILARITY).
FT   CHAIN         86   1108       COLLAGEN ALPHA 2(I) CHAIN.
FT   PROPEP     1109   1372       CARBOXYL-TERMINAL PROPEPTIDE
FT                                (BY SIMILARITY).
FT   MOD_RES       86     86       PYRROLIDONE CARBOXYLIC ACID (BY
FT                                SIMILARITY).
FT   MOD_RES       90     90       CONVERTED TO AN ALDEHYDE GROUP THAT IS
```

```
FT                                INVOLVED IN CROSS-LINKING
FT                                (BY SIMILARITY).
FT   CARBOHYD    1273   1273      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      15     15      V -> A (IN REF. 4).
FT   CONFLICT    1167   1167      R -> TT (IN REF. 1).
SQ   SEQUENCE   1372 AA;  129557 MW;  0D17DF5D6C1452D1 CRC64;


  Query Match             61.3%;  Score 204;  DB 1;  Length 1372;
  Best Local Similarity   66.7%;  Pred. No. 1.9e-09;
  Matches   38;  Conservative    4;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKG 57
           | || | :|| |||| : || || || |  | |||||| || | ||:|| :|||||||
Db     449 EPGLMGPRGLPGSPGNVGPSGKEGPVGLPGIDGRPGPIGPAGPRGEAGNIGFPGPKG 505


Search completed: January 28, 2002, 07:48:32
Job time: 97 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:46:55 ; Search time 37.99 Seconds
(without alignments)
227.167 Million cell updates/sec

Title: US-09-710-239-18
Perfect score: 333
Sequence: 1 EAGLPGAKGLTGSPGSPGPD.....PPGARGQAGVMGFPGPKGAA 59

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	333	100.0	138	4	Q9UML6	Q9um16 homo sapien
2	333	100.0	589	4	Q13896	Q13896 homo sapien
3	333	100.0	1461	4	O76045	O76045 homo sapien
4	321	96.4	1453	11	Q63079	Q63079 rattus norv
5	304	91.3	1445	13	O93251	O93251 rana catesb
6	296	88.9	1450	13	Q9YIB4	Q9yib4 cynops pyrr
7	288	86.5	1447	13	Q9IB91	Q9ib91 xenopus lae
8	261	78.4	1491	13	Q91718	Q91718 xenopus lae
9	258	77.5	1486	13	Q91717	Q91717 xenopus lae
10	257	77.2	1418	13	Q9W7R9	Q9w7r9 cynops pyrr
11	256	76.9	1160	4	Q14046	Q14046 homo sapien
12	256	76.9	1418	6	Q28396	Q28396 equus cabal
13	256	76.9	1419	11	Q63123	Q63123 rattus norv
14	256	76.9	1442	11	Q62031	Q62031 mus musculu
15	256	76.9	1442	11	Q62033	Q62033 mus musculu
16	256	76.9	1459	11	Q62032	Q62032 mus musculu
17	256	76.9	1487	4	Q14047	Q14047 homo sapien
18	256	76.9	1487	6	O77753	O77753 canis famil
19	232	69.7	886	13	Q92029	Q92029 gallus gall
20	225	67.6	1497	11	Q61431	Q61431 mus musculu
21	216	64.9	940	13	O93484	O93484 oncorhynchu
22	202	60.7	1372	11	Q9R1E8	Q9r1e8 rattus norv
23	201	60.4	1186	4	Q9UEB6	Q9ueb6 homo sapien
24	201	60.4	1366	4	Q15177	Q15177 homo sapien
25	200	60.1	1366	4	Q9UPH0	Q9uph0 homo sapien
26	193	58.0	1355	13	O42350	O42350 rana catesb
27	187.5	56.3	1691	11	Q9ESQ2	Q9esq2 mus musculu
28	183	55.0	301	5	Q19763	Q19763 caenorhabdi
29	183	55.0	771	4	Q9UJC7	Q9ujc7 homo sapien
30	180	54.1	675	13	Q90800	Q90800 gallus gall
31	178.5	53.6	890	5	O77087	O77087 alvinella p
32	178	53.5	303	5	Q93208	Q93208 caenorhabdi
33	177.5	53.3	1835	13	Q9IAU4	Q9iau4 gallus gall
34	176.5	53.0	302	5	Q19079	Q19079 caenorhabdi
35	176.5	53.0	1621	4	Q9H4R9	Q9h4r9 homo sapien
36	176.5	53.0	1669	11	Q9QZS0	Q9qzs0 mus musculu
37	176	52.9	1414	5	Q26634	Q26634 strongyloce
38	175	52.6	296	5	Q22389	Q22389 caenorhabdi
39	174.5	52.4	142	6	Q9BDZ1	Q9bdz1 macaca mula
40	174.5	52.4	632	5	Q9N2N7	Q9n2n7 hemicentrot
41	174	52.3	622	4	Q9BY85	Q9by85 homo sapien
42	174	52.3	742	4	Q9BYH7	Q9byh7 homo sapien
43	173.5	52.1	452	5	Q17189	Q17189 brugia mala
44	173.5	52.1	886	4	Q9NUB7	Q9nub7 homo sapien
45	173.5	52.1	1140	11	Q61434	Q61434 mus musculu

ALIGNMENTS

RESULT 1
Q9UML6
ID Q9UML6 PRELIMINARY; PRT; 138 AA.
AC Q9UML6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ALPHA-1 TYPE I COLLAGEN (FRAGMENT).
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85190598; PubMed=3857621;
RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
RT "Intron-mediated recombination may cause a deletion in an alpha 1 type
RT I collagen chain in a lethal form of osteogenesis imperfecta.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
DR EMBL; M11162; AAA75386.1; -.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 2.
KW Collagen.
FT NON_TER 1 1
SQ SEQUENCE 138 AA; 12129 MW; 34CEF270C29F7A7B CRC64;

Query Match 100.0%; Score 333; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.1e-26;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 44 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 102

RESULT 2
Q13896
ID Q13896 PRELIMINARY; PRT; 589 AA.
AC Q13896;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)

```
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-1 TYPE I COLLAGEN PRECURSOR (FRAGMENT).
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   D'Alessio M.;
RL   Submitted (FEB-1989) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88329734; PubMed=2843432;
RA   D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.,
RA   Pretorious P.J.;
RT   "Complete nucleotide sequence of the region encompassing the first
RT   twenty-five exons of the human pro alpha 1(I) collagen gene
RT   (COL1A1).";
RL   Gene 67:105-115(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Marini J.C., Lewis M.B., Wang Q., Chen K.C., Orrison B.M.;
RL   J. Biol. Chem. 0:0-0(0).
DR   EMBL; M20789; AAB59373.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 7.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00214; VWC; 1.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN       179   >589       ALPHA-1 TYPE I COLLAGEN.
FT   VARIANT     353    353       G -> S (IN REF. 3).
FT   NON_TER     589    589
SQ   SEQUENCE   589 AA;  55060 MW;  4148B73699BB9C4B CRC64;


  Query Match             100.0%;  Score 333;  DB 4;  Length 589;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-26;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 589


RESULT   3
O76045
ID   O76045      PRELIMINARY;   PRT;   1461 AA.
AC   O76045;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PRO ALPHA 1(I) COLLAGEN.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85130970; PubMed=2857713;
RA   Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT   Promoter structure, AluI repeats, and polymorphic transcripts.";
RL   J. Biol. Chem. 260:2315-2320(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88329734; PubMed=2843432;
RA   D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT   "Complete nucleotide sequence of the region encompassing the first
RT   twenty-five exons of the human pro alpha 1(I) collagen gene
RT   (COL1A1).";
RL   Gene 67:105-115(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89025644; PubMed=3178743;
RA   Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA   Jaenisch R., Prockop D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT   of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91138770; PubMed=1995349;
RA   Maatta A., Bornstein P., Penttinen R.P.;
RT   "Highly conserved sequences in the 3'-untranslated region of the
RT   COL1A1 gene bind cell-specific nuclear proteins.";
RL   FEBS Lett. 279:9-13(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92157916; PubMed=1787829;
RA   Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA   Olsen A.S., Prockop D.J.;
RT   "Completion of the last half of the structure of the human gene for
RT   the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL   Matrix 11:375-379(1991).
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Korkko J.M., Earley J.J., Nuytinck L., DePaepe A., Prockop D.J.,
RA   Ala-Kokko L.;
RT   "Analysis of the COL1A1 and COL1A2 Genes by CSGE and DNA Sequencing in
RT   12 Patients with mild OI (Type I). Identification of Common Sequences
RT   for Null Allele Mutations.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF017178; AAB94054.2; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
KW   Collagen.
SQ   SEQUENCE   1461 AA;  138629 MW;  9ACF6DE30EA78E21 CRC64;


  Query Match             100.0%;  Score 333;  DB 4;  Length 1461;
  Best Local Similarity   100.0%;  Pred. No. 1e-25;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    528 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 586


RESULT   4
Q63079
ID   Q63079      PRELIMINARY;      PRT;  1453 AA.
AC   Q63079;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COLLAGEN ALPHA1 (FRAGMENT).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1092 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=TOOTH;
```

```
RA   Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z78279; CAB01633.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   1453 AA;  137886 MW;  E6896BDC19A4A1D8 CRC64;


  Query Match             96.4%;  Score 321;  DB 11;  Length 1453;
  Best Local Similarity   96.6%;  Pred. No. 1.7e-24;
  Matches   57;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||| |
Db    520 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGARGQAGVMGFPGPKGTA 578


RESULT   5
O93251
ID   O93251      PRELIMINARY;      PRT;  1445 AA.
AC   O93251;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA 1 TYPE I COLLAGEN.
OS   Rana catesbeiana (Bull frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX   NCBI_TaxID=8400;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Asahina K., Utoh R., Obara M., Yoshizato K.;
RT   "Spatiotemporal expression of bullfrog a1(I) and a2(I) collagen genes
RT   in intestine during metamorphosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB015440; BAA29028.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
SQ   SEQUENCE   1445 AA;  137251 MW;  F59BB550C9873F04 CRC64;


  Query Match             91.3%;  Score 304;  DB 13;  Length 1445;
  Best Local Similarity   91.5%;  Pred. No. 8.5e-23;
  Matches   54;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          | ||||||||||||||||||| |  ||||||||||||||||:||||||||||||
Db    516 EPGLPGAKGLTGSPGSPGPDGKTGPAGAPGQDGRPGPPGPPGARGQSGVMGFPGPKGAA 574


RESULT   6
Q9YIB4
ID   Q9YIB4      PRELIMINARY;      PRT;  1450 AA.
AC   Q9YIB4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA 1 TYPE I COLLAGEN.
OS   Cynops pyrrhogaster (Japanese common newt).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX   NCBI_TaxID=8330;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=REGENERATE FORELIMBS;
RA   Asahina K., Obarà M., Yoshizato K.;
RT   "Cynops pyrrhogaster alpha 1 type I collagen, partial cDNA.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB015438; BAA36973.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; UNKNOWN_1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
KW   Collagen.
SQ   SEQUENCE   1450 AA;  137563 MW;  ABF8A74841B87B7C CRC64;


  Query Match             88.9%;  Score 296;  DB 13;  Length 1450;
  Best Local Similarity   89.8%;  Pred. No. 5.5e-22;
  Matches   53;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          | |||||||||||||||||||| | ||||| |||||| |||||:||||||||||||
Db    517 EPGLPGAKGLTGSPGSPGPDGKTGPAGAAGQDGHPGPPGPSGARGQSGVMGFPGPKGAA 575


RESULT   7
Q9IB91
ID   Q9IB91      PRELIMINARY;      PRT;  1447 AA.
AC   Q9IB91;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TYPE I COLLAGEN ALPHA 1.
GN   COL1A1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RT   "Expression and characterization of Xenopus type I collagen alpha 1
RT   (COL1A1) during embryonic development.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034701; BAA94972.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Collagen.
SQ   SEQUENCE   1447 AA;  137445 MW;  A4A6DD2B4158E38B CRC64;


  Query Match             86.5%;  Score 288;  DB 13;  Length 1447;
  Best Local Similarity   88.1%;  Pred. No. 3.5e-21;
  Matches   52;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAA 59
          | ||||||||||||||| |||||| |  ||||| ||||||||||:||||||||||||
Db    516 EPGLPGAKGLTGSPGSPGSDGKTGPAGAPGQDGRAGPPGPPGARGQSGVMGFPGPKGAA 574
```

```
RESULT   8
Q91718
ID   Q91718      PRELIMINARY;      PRT;  1491 AA.
AC   Q91718;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-1 TYPE II' COLLAGEN.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92011898; PubMed=1918153;
RA   Su M., Suzuki H.R., Bieker J.J., Solursh M., Ramirez F.;
RT   "Expression of two nonallelic type II procollagen genes during Xenopus
RT   laevis embryogenesis is characterized by stage-specific production of
RT   alternatively spliced transcripts.";
RL   J. Cell Biol. 115:565-575(1991).
DR   EMBL; M63596; AAA49679.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
SQ   SEQUENCE   1491 AA;  142495 MW;  43026FFF08FB0314 CRC64;


  Query Match             78.4%;  Score 261;  DB 13;  Length 1491;
  Best Local Similarity   79.3%;  Pred. No. 1.9e-18;
  Matches   46;  Conservative    2;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | |||||:|||| ||  || || || | ||:|||||||||| ||||| ||||||||||
Db     558 EPGLPGARGLTGRPGDAGPQGKVGPSGAAGEDGRPGPPGPQGARGQPGVMGFPGPKGA 615


RESULT   9
Q91717
ID   Q91717      PRELIMINARY;      PRT;  1486 AA.
AC   Q91717;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-1 TYPE II COLLAGEN.
GN   COL2A1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92011898; PubMed=1918153;
RA   Su M., Suzuki H.R., Bieker J.J., Solursh M., Ramirez F.;
RT   "Expression of two nonallelic type II procollagen genes during Xenopus
RT   laevis embryogenesis is characterized by stage-specific production of
RT   alternatively spliced transcripts.";
RL   J. Cell Biol. 115:565-575(1991).
DR   EMBL; M63595; AAA49678.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
SQ   SEQUENCE   1486 AA;  142263 MW;  4A4A95772341042F CRC64;


  Query Match             77.5%;  Score 258;  DB 13;  Length 1486;
  Best Local Similarity   77.6%;  Pred. No. 3.7e-18;
  Matches   45;  Conservative    3;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | |||||:|||| ||  || || || | :|:|||||||||| ||||| ||||||||||
Db     555 EPGLPGARGLTGRPGDAGPQGKVGPSGASGEDGRPGPPGPQGARGQPGVMGFPGPKGA 612


RESULT  10
Q9W7R9
ID   Q9W7R9      PRELIMINARY;      PRT;  1418 AA.
AC   Q9W7R9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA1 TYPE II COLLAGEN.
OS   Cynops pyrrhogaster (Japanese common newt).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX   NCBI_TaxID=8330;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Asahina K., Obara M., Yoshizato K.;
RT   "Unique expression of genes of type I and type II collagens of
RT   regenerating newt limb in apical epidermal cap, blastema, muscle and
RT   cartilage.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB022046; BAA82043.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Collagen.
SQ   SEQUENCE   1418 AA;  135066 MW;  C19A6E601A2A717E CRC64;


  Query Match             77.2%;  Score 257;  DB 13;  Length 1418;
  Best Local Similarity   77.6%;  Pred. No. 4.5e-18;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
           | |||| :|||| ||  || || || | ||:|||||||||| ||||| ||||||||||
Db     484 EPGLPGVRGLTGHPGDAGPQGKVGPTGAAGEDGRPGPPGPQGARGQPGVMGFPGPKGA 541


RESULT  11
Q14046
ID   Q14046      PRELIMINARY;      PRT;  1160 AA.
AC   Q14046;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COL2A1 ALPHA1 (II) COLLAGEN PRECURSOR (FRAGMENT).
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CARTILAGE;
```

```
RX   MEDLINE=90026318; PubMed=2803268;
RA   Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
RT   "Structure of cDNA clones coding for human type II procollagen. The
RT   alpha 1(II) chain is more similar to the alpha 1(I) chain than two
RT   other alpha chains of fibrillar collagens.";
RL   Biochem. J. 262:521-528(1989).
DR   EMBL; X16711; CAA34683.1; -.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 18.
KW   Signal; Matrix protein.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN       113  >1160       COLLAGEN.
FT   NON_TER    1160   1160
SQ   SEQUENCE   1160 AA;  105630 MW;  A7F0523B856C8639 CRC64;


  Query Match             76.9%;  Score 256;  DB 4;  Length 1160;
  Best Local Similarity   77.6%;  Pred. No. 4.7e-18;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || |  |:|||||||||| ||||| ||||||||||
Db    484 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 541


RESULT  12
Q28396
ID   Q28396      PRELIMINARY;      PRT;  1418 AA.
AC   Q28396;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TYPE II COLLAGEN.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Richardson D.W., Dodge G.R.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 18-68 FROM N.A.
RA   MacLeod J.N., Fubini S.L., Gu D.N., Tetreault J.W., Todhunter R.J.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U62528; AAB05773.1; -.
DR   EMBL; AF040638; AAB96768.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
SQ   SEQUENCE   1418 AA;  134343 MW;  115FCD19EB8696A3 CRC64;


  Query Match             76.9%;  Score 256;  DB 6;  Length 1418;
  Best Local Similarity   77.6%;  Pred. No. 5.7e-18;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || |  |:|||||||||| ||||| ||||||||||
Db    484 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 541


RESULT  13
Q63123
ID   Q63123      PRELIMINARY;      PRT;  1419 AA.
AC   Q63123; Q63565;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COLLAGEN ALPHA 1 TYPE II (T1 MRNA).
GN   T1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE FRACTURE CALLUS;
RA   Urabe K., Sarkar G., Bolander M.E.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1372-1419 FROM N.A.
RA   Wurtz T., Brandsten C., Lumdmark C., Christersson C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 370-422 FROM N.A.
RC   STRAIN=DA; TISSUE=CARTILLAGE;
RX   MEDLINE=94321934; PubMed=8046350;
RA   Michaelson E., Malmstrom V., Reis S., Engstrom A., Burkhardt H.,
RA   Holmdahl R.;
RT   "T cell recognition of carbohydrates on type II collagen.";
RL   J. Exp. Med. 180:745-749(1994).
DR   EMBL; L48440; AAA79780.1; -.
DR   EMBL; AJ224879; CAA12179.1; -.
DR   EMBL; X79816; CAA56213.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
SQ   SEQUENCE   1419 AA;  134570 MW;  B7C63B77819CE50B CRC64;


  Query Match             76.9%;  Score 256;  DB 11;  Length 1419;
  Best Local Similarity   77.6%;  Pred. No. 5.7e-18;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || |  |:|||||||||| ||||| ||||||||||
Db    485 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 542


RESULT  14
Q62031
ID   Q62031      PRELIMINARY;      PRT;  1442 AA.
AC   Q62031;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PRO-ALPHA-1 TYPE II COLLAGEN.
GN   COL2A1 OR PRO-ALPHA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RX   MEDLINE=91358489; PubMed=1885613;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. Biol. Chem. 266:16862-16869(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (OCT-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
```

```
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M65161; AAA68099.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   1442 AA;  137911 MW;  D3A3274493CB821C CRC64;


  Query Match           76.9%;  Score 256;  DB 11;  Length 1442;
  Best Local Similarity 77.6%;  Pred. No. 5.8e-18;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || |  |:|||||||| ||||| ||||||||||
Db    508 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 565


RESULT  15
Q62033
ID   Q62033      PRELIMINARY;      PRT;  1442 AA.
AC   Q62033;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PRO-ALPHA-1 TYPE II COLLAGEN.
GN   COL2A1 OR PRO-ALPHA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RX   MEDLINE=91358489; PubMed=1885613;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. Biol. Chem. 266:16862-16869(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (OCT-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M65161; AAA68102.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   1442 AA;  137828 MW;  F0E77C11BCAFA93B CRC64;


  Query Match           76.9%;  Score 256;  DB 11;  Length 1442;
  Best Local Similarity 77.6%;  Pred. No. 5.8e-18;
  Matches   45;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 EAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGA 58
          | |||||:|||| ||  || || || |  |:|||||||| ||||| ||||||||||
Db    508 EPGLPGARGLTGRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGA 565


Search completed: January 28, 2002, 07:48:11
Job time: 76 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:48:59 ; Search time 19.73 Seconds
(without alignments)
114.056 Million cell updates/sec

Title: US-09-710-239-29
Perfect score: 580
Sequence: 1 RGDKGETGEQGDRGIKGHRG.....DAGPVGPPGPPGPPGPPGPP 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	580	100.0	1057	3	US-08-931-820-1	Sequence 1, Appli
2	580	100.0	1341	3	US-08-963-825-18	Sequence 18, Appl
3	472	81.4	1442	2	US-08-316-650-12	Sequence 12, Appl
4	472	81.4	1442	5	PCT-US95-02251-12	Sequence 12, Appl
5	461	79.5	1060	3	US-08-931-820-3	Sequence 3, Appli
6	461	79.5	1418	3	US-08-963-825-20	Sequence 20, Appl
7	453	78.1	1418	4	US-09-010-999-1	Sequence 1, Appli
8	392	67.6	1057	3	US-08-931-820-4	Sequence 4, Appli
9	384	66.2	1078	3	US-08-963-825-21	Sequence 21, Appl
10	348	60.0	535	4	US-09-029-348-1	Sequence 1, Appli
11	348	60.0	537	4	US-09-029-348-4	Sequence 4, Appli
12	348	60.0	1366	3	US-08-963-825-19	Sequence 19, Appl
13	328	56.6	534	4	US-09-029-348-5	Sequence 5, Appli
14	328	56.6	1024	3	US-08-931-820-2	Sequence 2, Appli
15	315.5	54.4	357	1	US-07-609-716-66	Sequence 66, Appl
16	315.5	54.4	357	1	US-08-642-255-33	Sequence 33, Appl
17	315.5	54.4	357	4	US-08-475-411A-66	Sequence 66, Appl
18	315.5	54.4	357	4	US-08-478-029A-66	Sequence 66, Appl
19	315	54.3	330	1	US-08-642-255-32	Sequence 32, Appl
20	315	54.3	408	1	US-07-609-716-65	Sequence 65, Appl
21	315	54.3	408	4	US-08-475-411A-65	Sequence 65, Appl
22	315	54.3	408	4	US-08-478-029A-65	Sequence 65, Appl
23	306	52.8	504	4	US-09-219-849-3	Sequence 3, Appli
24	306	52.8	561	1	US-08-642-255-52	Sequence 52, Appl
25	306	52.8	720	4	US-09-219-849-4	Sequence 4, Appli
26	306	52.8	777	1	US-08-642-255-53	Sequence 53, Appl
27	305	52.6	144	1	US-08-642-255-49	Sequence 49, Appl
28	297	51.2	532	1	US-08-494-168-9	Sequence 9, Appli
29	296	51.0	234	1	US-08-642-255-51	Sequence 51, Appl
30	295	50.9	822	4	US-09-219-849-49	Sequence 49, Appl
31	287	49.5	595	4	US-09-219-849-48	Sequence 48, Appl
32	287	49.5	595	4	US-09-219-849-50	Sequence 50, Appl
33	286.5	49.4	546	1	US-08-494-168-10	Sequence 10, Appl
34	285.5	49.2	1064	1	US-08-642-255-62	Sequence 62, Appl
35	283.5	48.9	252	1	US-08-642-255-61	Sequence 61, Appl
36	280	48.3	471	2	US-08-399-889-24	Sequence 24, Appl
37	280	48.3	471	3	US-09-167-364-24	Sequence 24, Appl
38	280	48.3	471	4	US-09-439-897-2	Sequence 2, Appli
39	278.5	48.0	279	4	US-09-010-999-2	Sequence 2, Appli
40	278.5	48.0	310	4	US-09-219-849-47	Sequence 47, Appl
41	277.5	47.8	684	1	US-08-555-669-12	Sequence 12, Appl
42	277.5	47.8	684	3	US-09-073-663-12	Sequence 12, Appl
43	273.5	47.2	446	2	US-08-836-854-15	Sequence 15, Appl
44	273	47.1	960	4	US-09-219-849-5	Sequence 5, Appli
45	271	46.7	464	2	US-08-836-854-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT   1
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Assay for collagen degradation
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 96202596.1
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1057 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Collagen type I
US-08-931-820-1


  Query Match             100.0%;  Score 580;  DB 3;  Length 1057;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-37;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     932 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 991

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||||
Db     992 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 1031


RESULT   2
US-08-963-825-18
```

```
; Sequence 18, Application US/08963825
; Patent No. 6110689
;  GENERAL INFORMATION:
;    APPLICANT:  Qvist, Per
;    APPLICANT:  Bonde, Martin
;    TITLE OF INVENTION:  A Method for Assaying Collagen Fragments
;    TITLE OF INVENTION:  in Body Fluids, A Test Kit and Means for Carrying Out the
;    TITLE OF INVENTION:  Method and Use of the Method to Diagnose the Presence of
;    TITLE OF INVENTION:  Disorders Associated with the Metabolism of
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/963,825
;      FILING DATE:
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,319
;      FILING DATE:  21-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda C
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  4305/08701
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1341 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  COLLAGEN ALPHA 1 (I)
US-08-963-825-18


  Query Match             100.0%;  Score 580;  DB 3;  Length 1341;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-37;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     969 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 1028

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||||
Db   1029 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 1068


RESULT   3
US-08-316-650-12
; Sequence 12, Application US/08316650
; Patent No. 5942496
;  GENERAL INFORMATION:
;    APPLICANT:  Bonadio, Jeffrey
;    APPLICANT:  Roessler, Blake J.
;    APPLICANT:  Goldstein, Steven A.
;    APPLICANT:  Lin, Wushan
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS
;    TITLE OF INVENTION:  FOR STIMULATING BONE CELLS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/316,650
;      FILING DATE:  30-SEP-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/199,780
;      FILING DATE:  30-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UMIC:008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1442 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-316-650-12


  Query Match             81.4%;  Score 472;  DB 2;  Length 1442;
  Best Local Similarity   78.0%;  Pred. No. 3.1e-29;
  Matches   78;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||:||||:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db    1070 RGDKGESGEQGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1129

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:||||||||||||:|: ||||||| ||||||||||
Db    1130 DGSNGIPGPIGPPGPRGRSGETGPVGPPGSPGPPGPPGPP 1169


RESULT   4
PCT-US95-02251-12
; Sequence 12, Application PC/TUS9502251
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR STIMULATING BONE
;    TITLE OF INVENTION:  CELLS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version
;      SOFTWARE:  #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02251
;      FILING DATE:  CONCURRENTLY HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/316,650
;      FILING DATE:  30-SEP-1994
;      CLASSIFICATION:
;      APPLICATION NUMBER:  US 08/199,780
;      FILING DATE:  18-FEB-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UMIC009P--
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1442 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US95-02251-12


  Query Match             81.4%;  Score 472;  DB 5;  Length 1442;
  Best Local Similarity   78.0%;  Pred. No. 3.1e-29;
  Matches   78;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          ||||||:||||:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db   1070 RGDKGESGEQGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1129

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          || ||:||||||||||||:|: ||||||| ||||||||||
Db   1130 DGSNGIPGPIGPPGPRGRSGETGPVGPPGSPGPPGPPGPP 1169


RESULT   5
US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Assay for collagen degradation
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 96202596.1
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Collagen type II
US-08-931-820-3


  Query Match             79.5%;  Score 461;  DB 3;  Length 1060;
  Best Local Similarity   76.0%;  Pred. No. 1.6e-28;
  Matches   76;  Conservative    9;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||| || |:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db    934 RGDKGEAGEPGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 993

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          || ||:||||||||||||:|: || |||| ||||||||||
Db    994 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP 1033


RESULT   6
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
;  GENERAL INFORMATION:
;    APPLICANT:  Qvist, Per
;    APPLICANT:  Bonde, Martin
;    TITLE OF INVENTION:  A Method for Assaying Collagen Fragments
;    TITLE OF INVENTION:  in Body Fluids, A Test Kit and Means for Carrying Out the
;    TITLE OF INVENTION:  Method and Use of the Method to Diagnose the Presence of
;    TITLE OF INVENTION:  Disorders Associated with the Metabolism of
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/963,825
;      FILING DATE:
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,319
;      FILING DATE:  21-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda C
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  4305/08701
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1418 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  COLLAGEN -ALPHA 1 (II)
US-08-963-825-20


  Query Match             79.5%;  Score 461;  DB 3;  Length 1418;
  Best Local Similarity   76.0%;  Pred. No. 2.1e-28;
  Matches   76;  Conservative    9;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
```

```
          |||||| || |:||:||||||:|||| ||||| |:|| || :||:|||||||  |  ||
Db   1046 RGDKGEAGEPGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1105

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPGPP 100
          || ||:|||||||||||||:|: || |||| ||||||||||
Db   1106 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPPGPP 1145


RESULT   7
US-09-010-999-1
; Sequence 1, Application US/09010999
; Patent No. 6132976
;  GENERAL INFORMATION:
;    APPLICANT:  Poole, Anthony R.
;    APPLICANT:  Hollander, Anthony P.
;    APPLICANT:  Billinghurst, R. C.
;    TITLE OF INVENTION:  IMMUNOASSAYS FOR THE MEASUREMENT OF
;    TITLE OF INVENTION:  COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/010,999
;      FILING DATE:  22-JAN-1998
;      CLASSIFICATION:  4335
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/448,501
;      FILING DATE:  17-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/984,123
;      FILING DATE:  04-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  032931/0212
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1418 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Human Type II Collagen
US-09-010-999-1


  Query Match             78.1%;  Score 453;  DB 4;  Length 1418;
  Best Local Similarity   75.0%;  Pred. No. 8.8e-28;
  Matches   75;  Conservative    9;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||| || |:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db   1046 RGDKGEAGEPGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1105

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          ||  |:||||||||||||:|: || |||| ||||||||||
Db   1106 DGAMGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP 1145
```

```
RESULT   8
US-08-931-820-4
; Sequence 4, Application US/08931820
; Patent No. 6010863
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Assay for collagen degradation
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 96202596.1
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1057 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Collagen type III
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1055
;      OTHER INFORMATION:  /label= Modified
;      OTHER INFORMATION:  /note= "Ala may be Pro"
US-08-931-820-4


  Query Match             67.6%;  Score 392;  DB 3;  Length 1057;
  Best Local Similarity   69.0%;  Pred. No. 3.3e-23;
  Matches   69;  Conservative    6;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||||||:|  |||||||| |  | || ||  |:||  |: ||||||| | :| |||
Db    943 RGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGPAGPRGPVGPSGPPGK 1002

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          || :| ||||||||||  |: |  | || || |||||||
Db   1003 DGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPP 1042


RESULT   9
US-08-963-825-21
; Sequence 21, Application US/08963825
; Patent No. 6110689
;  GENERAL INFORMATION:
;    APPLICANT:  Qvist, Per
;    APPLICANT:  Bonde, Martin
;    TITLE OF INVENTION:  A Method for Assaying Collagen Fragments
;    TITLE OF INVENTION:  in Body Fluids, A Test Kit and Means for Carrying Out the
;    TITLE OF INVENTION:  Method and Use of the Method to Diagnose the Presence of
;    TITLE OF INVENTION:  Disorders Associated with the Metabolism of
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/963,825
;       FILING DATE:
;       CLASSIFICATION:  436
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/187,319
;       FILING DATE:  21-JAN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gogoris, Adda C
;       REGISTRATION NUMBER:  29,714
;       REFERENCE/DOCKET NUMBER:  4305/08701
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-527-7700
;       TELEFAX:  212-753-6237
;       TELEX:  236687
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1078 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
;       CLONE:  COLLAGEN ALPHA 1 (III)
US-08-963-825-21


  Query Match             66.2%;  Score 384;  DB 3;  Length 1078;
  Best Local Similarity   68.0%;  Pred. No. 1.4e-22;
  Matches   68;  Conservative    6;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||||||:|  ||||||||  |   | || ||  |:||  |: ||||||| | :| |||
Db     944 RGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGPAGPRGPVGPSGPPGK 1003

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || :| ||||||||||  |: |  | || ||  ||||||
Db    1004 DGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQSGPPGPP 1043


RESULT  10
US-09-029-348-1
; Sequence 1, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
;  APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
;  TITLE OF INVENTION: NOVEL PROCOLLAGENS
;  FILE REFERENCE: d087857PUS LISTING
;  CURRENT APPLICATION NUMBER: US/09/029,348
;  CURRENT FILING DATE: 1998-05-07
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 535
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
;   OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-1


  Query Match             60.0%;  Score 348;  DB 4;  Length 535;
  Best Local Similarity   63.3%;  Pred. No. 4.3e-20;
  Matches   62;  Conservative    8;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| ||:| ||: | :| :|||| ||  |  |:||  |: |||||||| | :|  ||
Db     174 RGDKGEPGEKGPRGLPGFKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGK 233

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPG 98
           ||  | || :|| | ||  |  || ||||||||||||
Db     234 DGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPGPPG 271


RESULT  11
US-09-029-348-4
; Sequence 4, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
;  APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
;  TITLE OF INVENTION: NOVEL PROCOLLAGENS
;  FILE REFERENCE: d087857PUS LISTING
;  CURRENT APPLICATION NUMBER: US/09/029,348
;  CURRENT FILING DATE: 1998-05-07
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 537
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
;   OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-4


  Query Match             60.0%;  Score 348;  DB 4;  Length 537;
  Best Local Similarity   63.3%;  Pred. No. 4.4e-20;
  Matches   62;  Conservative    8;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| ||:| ||: | :| :|||| ||  |  |:||  |: |||||||| | :|  ||
Db     174 RGDKGEPGEKGPRGLPGFKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGK 233

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPG 98
           ||  | || :|| | ||  |  || ||||||||||||
Db     234 DGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPGPPG 271


RESULT  12
US-08-963-825-19
; Sequence 19, Application US/08963825
; Patent No. 6110689
;  GENERAL INFORMATION:
;     APPLICANT:  Qvist, Per
;     APPLICANT:  Bonde, Martin
;     TITLE OF INVENTION:  A Method for Assaying Collagen Fragments
;     TITLE OF INVENTION:  in Body Fluids, A Test Kit and Means for Carrying Out the
;     TITLE OF INVENTION:  Method and Use of the Method to Diagnose the Presence of
;     TITLE OF INVENTION:  Disorders Associated with the Metabolism of
;     NUMBER OF SEQUENCES:  21
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Darby & Darby PC
;       STREET:  805 Third Avenue
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/963,825
;       FILING DATE:
;       CLASSIFICATION:  436
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/187,319
;       FILING DATE:  21-JAN-1994
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda C
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  4305/08701
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1366 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  collagen alpha 2- type I
US-08-963-825-19


  Query Match             60.0%;  Score 348;  DB 3;  Length 1366;
  Best Local Similarity   63.3%;  Pred. No. 1e-19;
  Matches   62;  Conservative    8;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||| ||:| ||: | :| :|||| ||  |  |:||  |: |||||||| | :|  ||
Db   1005 RGDKGEPGEKGPRGLPGFKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGK 1064

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPG 98
          ||  | || :|| | ||  |  || ||||||||||||||
Db   1065 DGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPGPPG 1102


RESULT  13
US-09-029-348-5
; Sequence 5, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
;  APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
;  TITLE OF INVENTION: NOVEL PROCOLLAGENS
;  FILE REFERENCE: d087857PUS LISTING
;  CURRENT APPLICATION NUMBER: US/09/029,348
;  CURRENT FILING DATE: 1998-05-07
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 534
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
;   OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-5


  Query Match             56.6%;  Score 328;  DB 4;  Length 534;
  Best Local Similarity   61.2%;  Pred. No. 1.5e-18;
  Matches   60;  Conservative    8;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||| ||:| ||: | :| :|||| ||  |  |:||  |: |||||||| | :|  ||
Db    173 RGDKGEPGEKGPRGLPGFKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGK 232

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPG 98
          ||  | || :|| | ||  |  || ||||||||| || |
Db    233 DGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPLGPLG 270


RESULT  14
US-08-931-820-2
; Sequence 2, Application US/08931820
; Patent No. 6010863
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Assay for collagen degradation
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 96202596.1
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1024 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Collagen type I
US-08-931-820-2


  Query Match             56.6%;  Score 328;  DB 3;  Length 1024;
  Best Local Similarity   61.2%;  Pred. No. 2.7e-18;
  Matches   60;  Conservative    8;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||| ||:| ||: | :| :|||| ||  |  |:||  |: |||||||| | :|  ||
Db    926 RGDKGEPGEKGPRGLPGFKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGK 985

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPG 98
          ||  | || :|| | ||  |  || ||||||||| || |
Db    986 DGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPLGPLG 1023


RESULT  15
US-07-609-716-66
; Sequence 66, Application US/07609716
; Patent No. 5514581
;  GENERAL INFORMATION:
;    APPLICANT:  Ferrari, Franco A.
;    APPLICANT:  Cappello, Joseph
;    TITLE OF INVENTION:  Functional Recombinantly Prepared
;    TITLE OF INVENTION:  Synthetic Protein Polymer
;    NUMBER OF SEQUENCES:  118
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/609,716
;      FILING DATE:  06-NOV-1990
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20015
;      REFERENCE/DOCKET NUMBER:  A-55186-3/BIR
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  357 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-609-716-66


  Query Match             54.4%;  Score 315.5;  DB 1;  Length 357;
  Best Local Similarity   52.2%;  Pred. No. 9.5e-18;
  Matches   60;  Conservative    5;  Mismatches   35;  Indels   15;  Gaps    2;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAP-- 58
           :||:|: | :|  |  |  |  |  | |||||  ||  ||  ||  ||  ||  ||||  | |
Db      69 KGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGLPGP 128

Qy      59 -------GKDGLNGLPGPIGP------PGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
                  |  | :| |||  ||      |||  |  |  ||  | ||||||||||||
Db     129 KGDRGDAGPKGADGSPGPAGPVGSPGAPGPPGPPGPPGPPGAPGPPGPPGPPGPP 183


Search completed: January 28, 2002, 07:49:00
Job time: 125 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:47:26 ; Search time 21.88 Seconds
(without alignments)
348.147 Million cell updates/sec

Title: US-09-710-239-29
Perfect score: 580
Sequence: 1 RGDKGETGEQGDRGIKGHRG.....DAGPVGPPGPPGPPGPPGPP 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	580	100.0	1464	1	CGHU1S	collagen alpha 1(I
2	565	97.4	779	1	CGBO1S	collagen alpha 1(I
3	562	96.9	1453	2	S21626	collagen alpha 1(I
4	558	96.2	473	2	I50629	collagen - chicken
5	558	96.2	1042	1	CGCH1S	collagen alpha 1(I
6	483	83.3	671	1	CGRT1S	collagen alpha 1(I
7	472	81.4	1419	2	A41182	collagen alpha 1(I
8	472	81.4	1487	2	B41182	collagen alpha 1(I
9	461	79.5	1418	2	T45467	collagen alpha 1(I
10	461	79.5	1487	1	CGHU6C	collagen alpha 1(I
11	451	77.8	1486	1	B40333	collagen alpha 1(I
12	449	77.4	1492	2	A40333	collagen alpha 1'(
13	447	77.1	464	2	S59513	collagen II A1 pro
14	439	75.7	193	2	S07133	collagen alpha 1(I
15	410	70.7	1497	2	I49607	procollagen type V
16	404	69.7	1496	1	CGHU2V	collagen alpha 2(V
17	400	69.0	1464	2	S59856	collagen alpha 1(I
18	399	68.8	636	2	S41067	collagen alpha 1(I
19	395	68.1	964	1	CGCH2S	collagen alpha 2(I
20	392	67.6	1466	1	CGHU7L	collagen alpha 1(I
21	390	67.2	1049	1	CGBO7S	collagen alpha 1(I
22	376	64.8	365	2	S10847	collagen alpha 2(I
23	359	61.9	1373	1	A43291	collagen alpha 2(I
24	348	60.0	1366	1	CGHU2S	collagen alpha 2(I
25	342.5	59.1	615	2	A05269	collagen alpha 1(I
26	308.5	53.2	1843	2	S18803	collagen alpha 1(V
27	307.5	53.0	1838	1	CGHU1V	collagen alpha 1(V
28	306.5	52.8	1146	2	A38587	collagen, cornea-s
29	302	52.1	1027	2	S28774	collagen alpha cha
30	300.5	51.8	888	2	S28791	collagen alpha 1(X
31	298	51.4	675	2	S20819	collagen alpha 3(I
32	298	51.4	1315	2	A56101	collagen alpha 1(X
33	298	51.4	1774	2	B56101	collagen alpha 1(X
34	297	51.2	1669	1	CGHU4B	collagen alpha 1(I
35	296.5	51.1	1806	1	CGHU1E	collagen alpha 1(X
36	295.5	50.9	1669	1	CGMS4B	collagen alpha 1(I
37	295	50.9	1024	2	S18251	collagen alpha 1(X
38	294	50.7	632	2	S42731	collagen alpha 1 c
39	293.5	50.6	366	2	S11449	collagen short cha
40	292.5	50.4	469	2	A24450	collagen alpha 2(V
41	292.5	50.4	1546	1	CGHU2E	collagen alpha 2(X
42	291	50.2	730	2	A36226	collagen alpha 1 c
43	291	50.2	1414	1	S23809	collagen alpha 2(I
44	290	50.0	674	2	S23297	collagen alpha 1(X
45	290	50.0	3124	1	A40020	collagen alpha 1(X

ALIGNMENTS

RESULT 1
CGHU1S
collagen alpha 1(I) chain precursor - human
N;Alternate names: procollagen alpha 1(I) chain
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 31-Dec-2000
C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; 5269; A29439; I53466; A02852; I37247
R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorious, P
Gene 67, 105-115, 1988
A;Title: Complete nucleotide sequence of the region encompassing the first twenty-fiv
A;Reference number: I60114; MUID:88329734
A;Accession: I60114
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-369,'L',371-589 <DAL>
A;Cross-references: GB:M20789; NID:g179593; PIDN:AAB59373.1; PID:g179594
R;Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Pr
Biochem. J. 253, 919-922, 1988
A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of hum
A;Reference number: S01143; MUID:89025644
A;Accession: S01143
A;Molecule type: mRNA
A;Residues: 1-472 <TRO>
A;Cross-references: EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID:g30016; GB:M36546;
A;Note: submitted to the EMBL/GenBank/DDBJ databases by Prockop, D.J., 13-JUN-1988
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams,
Nature 310, 337-340, 1984
A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation
A;Reference number: A93335; MUID:84270697
A;Accession: A93335
A;Molecule type: DNA
A;Residues: 1-58,'Q',60-181 <CHU>
A;Cross-references: EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID:g35658
R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet,
J. Biol. Chem. 262, 15151-15157, 1987
A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene
A;Reference number: I55254; MUID:88033098
A;Accession: I55254
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-45 <ROS>
A;Cross-references: GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:g180388
R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A;Title: Regulatory elements in the first intron contribute to transcriptional contro
A;Reference number: A39943; MUID:88097389
A;Accession: A39943
A;Molecule type: DNA
A;Residues: 1-34 <BOR>
A;Cross-references: GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:g553238
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A;Reference number: I55237; MUID:85130970
A;Accession: I55237
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-34 <CH2>
A;Cross-references: GB:M10627; NID:g180383; PIDN:AAA51992.1; PID:g553226
R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A;Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
rome, type VII.
A;Reference number: A35233; MUID:90202908
A;Accession: A35233
A;Molecule type: protein
A;Residues: 33-52 <WIR>
A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A;Title: A base substitution in the exon of a collagen gene causes alternative splicing
A;Reference number: S09400; MUID:89356643
A;Accession: S09400
A;Molecule type: mRNA
A;Residues: 156-183 <WEI>
R;Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A;Reference number: A90567; MUID:71038625
A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A;Accession: B90567
A;Molecule type: protein
A;Residues: 162-198,'Z',200-201,'Z',203-206,'Z',208-209,'Z',211-228,'B',230,'BB',233,'Z'
A;Experimental source: skin
A;Note: evidence for 170-allysine
R;Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, P
Eur. J. Biochem. 192, 153-159, 1990
A;Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A;Reference number: S11372; MUID:90382436
A;Accession: S11372
A;Molecule type: protein
A;Residues: 175-187;274-287,'P',289 <BAE>
A;Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A;Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
ooperative melting of intact type I collagen.
A;Reference number: I55342; MUID:92042092
A;Accession: I55342
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 258-268;1347-1357 <DEA>
A;Cross-references: GB:S67495; NID:g239007; PIDN:AAB20350.1; PID:g239008
A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A;Reference number: A92069; MUID:71001508
A;Accession: A92069
A;Molecule type: protein
A;Residues: 263-268 <MOR>
A;Experimental source: skin
A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R;Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A;Title: Segmental amplification of the entire helical and telopeptide regions of the cD
A;Reference number: S15989; MUID:90326017
A;Accession: S15989
A;Molecule type: mRNA
A;Residues: 281-302;402-420;823-843;925-944;1026-1045;1143-1162 <LAB>
R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
Connect. Tissue Res. 29, 1-11, 1993
A;Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A;Reference number: I52905; MUID:93339042
A;Accession: I52905
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 342-352,'C',354-359 <WI2>
A;Cross-references: GB:S64717; NID:g408195; PIDN:AAB27677.1; PID:g408196
A;Note: mutant sequence from patient with osteogenesis imperfecta
R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A;Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalph
A;Reference number: A90476; MUID:84080385
A;Accession: A90476
A;Molecule type: mRNA
A;Residues: 425-1250,'X',1252-1328,'S',1330-1390,'X',1392-1464 <BER>
A;Cross-references: GB:K01228; NID:g180391; PIDN:AAA51995.1; PID:g180392
A;Note: sequence partially completed for missing nucleotides by A29439
R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A;Title: Multiexon deletion in an osteogenesis imperfecta variant with increased type
A;Reference number: A22161; MUID:85104934
A;Accession: A22161
A;Molecule type: DNA
A;Residues: 472-594,'R',596-607 <CH3>
A;Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PIDN:AAA51847.1;
A;Note: the authors translated the codon CGT for residue 595 as Pro
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explai
.
A;Reference number: A35336; MUID:90252792
A;Accession: A35336
A;Molecule type: mRNA
A;Residues: 710-720,'E',722-737,'E',739-745 <WAL>
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mot
Hum. Mol. Genet. 3, 2201-2206, 1994
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in th
A;Reference number: I54365; MUID:95187161
A;Accession: I54365
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 746-766,'S',768-781 <FOR>
A;Cross-references: GB:L47667; NID:g1009093; PIDN:AAB59576.1; PID:g1009094
R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of
A;Reference number: A47426; MUID:93352646
A;Accession: A47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1336,1339-1387,'R',1389-1464 <CHE>
A;Cross-references: GB:S64596; NID:g407589; PIDN:AAB27856.1; PID:g407590
A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIP:136445)
A;Note: does not represent an experimentally determined sequence but three different
A;Accession: B47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH4>
A;Experimental source: normal dermal fibroblast culture
A;Accession: C47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1464 <CH5>
A;Experimental source: fetal cell 86-237
A;Accession: D47426
A;Molecule type: mRNA
A;Residues: 1179-1336,1339-1464 <CH6>
A;Experimental source: fetal cell 86-146
A;Accession: E47426
A;Molecule type: mRNA
A;Residues: 1179-1387,'R',1389-1464 <CH7>
A;Experimental source: fetal cell 88-251
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.;
J. Biol. Chem. 263, 14605-14607, 1988
A;Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptid
A;Reference number: I55269; MUID:89008319
A;Accession: I55269

```
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1187-1194,'C',1196-1220 <COH>
A;Cross-references: GB:M23213; NID:g340842; PIDN:AAB59363.1; PID:g499622
A;Note: mutant sequence from a patient with mild osteogenesis imperfecta
R;Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.
Nucleic Acids Res. 16, 349, 1988
A;Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

  Query Match            100.0%;  Score 580;  DB 1;  Length 1464;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-30;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1093 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 1152

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPGPP 100
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1153 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPGPP 1192


RESULT   2
CGBO1S
collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 31-Dec-1993 #text_change 31-Mar-2000
C;Accession: A91193; A91229; A91387; A91211; A91201; A91200; A43048; A02853
R;Rauterberg, J.; Timpl, R.; Furthmayr, H.
Eur. J. Biochem. 27, 231-237, 1972
A;Title: Structural characterization of N-terminal antigenic determinants in calf and hu
A;Reference number: A91193; MUID:72255334
A;Accession: A91193
A;Molecule type: protein
A;Residues: 1-19 <RAU>
A;Experimental source: skin
A;Note: the epsilon carbon of Lys-9, by homology with the rat alpha 1(I) chain, is conve
R;Fietzek, P.P.; Kuehn, K.
Eur. J. Biochem. 52, 77-82, 1975
A;Title: The covalent structure of collagen: amino-acid sequence of the cyanogen-bromide
A;Reference number: A91229; MUID:76022320
A;Accession: A91229
A;Molecule type: protein
A;Residues: 20-145 <FIE>
A;Experimental source: skin
A;Note: Lys-103 is hydroxylated and binds glucosylgalactose
R;Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.
FEBS Lett. 26, 74-76, 1972
A;Title: The covalent structure of collagen: amino acid sequence of alpha1-CB3 from calf
A;Reference number: A91387; MUID:73049499
A;Accession: A91387
A;Molecule type: protein
A;Residues: 146-294 <FI2>
A;Experimental source: skin
R;Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
Eur. J. Biochem. 38, 396-400, 1973
A;Title: The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 fr
A;Reference number: A91211; MUID:74086118
A;Accession: A91211
A;Molecule type: protein
A;Residues: 295-562 <FI3>
A;Experimental source: skin
R;Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
Eur. J. Biochem. 30, 169-183, 1972
A;Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
A;Reference number: A91201; MUID:73042276
A;Accession: A91201
A;Molecule type: protein
A;Residues: 563-675 <WEN>
A;Experimental source: skin
R;Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
Eur. J. Biochem. 30, 163-168, 1972
A;Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-d
A;Reference number: A91200; MUID:73042275
A;Accession: A91200
A;Molecule type: protein
A;Residues: 676-758 <FI4>
A;Experimental source: skin
A;Note: Pro-726 is the only 3-hydroxyproline and the only hydroxylated proline in pos
R;Rauterberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.
FEBS Lett. 21, 75-79, 1972
A;Title: The amino acid sequence of the carboxyterminal nonhelical cross link region
A;Reference number: A43048
A;Accession: A43048
A;Molecule type: protein
A;Residues: 759-779 <RA2>
A;Experimental source: skin
C;Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658, and 670 may be hyd
C;Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) ar
C;Comment: The order of the eight CNBr peptides in the alpha 1(I) chain of bovine ski
9, 149, 268, and 217 residues.
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trime
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

  Query Match            97.4%;  Score 565;  DB 1;  Length 779;
  Best Local Similarity  93.0%;  Pred. No. 8.8e-30;
  Matches   93;  Conservative    6;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          ||:||:||::|:||||||||||||||||||||||||||||||||||||||||||||:|||
Db    655 RGBKGZTGZZGBRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGSPGK 714

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          |||||||||||||||||||||| ||||||||||||||||
Db    715 DGLNGLPGPIGPPGPRGRTGDAGPAGPPGPPGPPGPPGPP 754


RESULT   3
S21626
collagen alpha 1(I) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 25-Apr-1997 #text_change 13-Aug-1999
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R;Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A;Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type
A;Reference number: S57243
A;Accession: S57243
A;Molecule type: mRNA
A;Residues: 1-1453 <LIS>
A;Cross-references: EMBL:U08020; NID:g470673; PIDN:AAA88912.1; PID:g470674
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNA
A;Reference number: S16176; MUID:91274355
A;Accession: S16374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1453 <MET>
A;Cross-references: EMBL:X57981; NID:g50484; PIDN:CAA41046.1; PID:g50485
R;French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A;Reference number: A23982; MUID:86137403
A;Accession: A23982
A;Molecule type: mRNA
A;Residues: 518-1128 <FRE>
A;Cross-references: GB:M14423; NID:g192261; PIDN:AAA37333.1; PID:g192262
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A;Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence f
A;Reference number: I49559; MUID:83141374
A;Accession: I49559
```

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 735-1130 <RES>
A;Cross-references: GB:M17491; NID:g192263; PIDN:AAA37334.1; PID:g192264
R;Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A;Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A;Reference number: I49557; MUID:84170331
A;Accession: I49557
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-25 <RE2>
A;Cross-references: GB:K01688; NID:g192246; PIDN:AAA37330.1; PID:g553881
R;Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A;Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A;Reference number: S39789; MUID:94092741
A;Accession: S39789
A;Molecule type: DNA
A;Residues: 1-80,'E',82-105,'D',107-185;1031-1201,'G',1203-1218,'E',1220-1221,'T',1223-1
R;Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A;Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirec
A;Reference number: I48300; MUID:94344105
A;Accession: I48300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-80,'E',82-105,'D',107-147 <REF>
A;Cross-references: EMBL:X54876; NID:g50486; PIDN:CAA38657.1; PID:g50487
C;Genetics:
A;Gene: COL1A1
A;Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F;30-89/Domain: von Willebrand factor type C repeat homology <VWC>
F;152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
F;1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             96.9%;  Score 562;  DB 2;  Length 1453;
  Best Local Similarity   96.0%;  Pred. No. 2.3e-29;
  Matches   96;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||:|||
Db    1082 RGDKGETGEQGDRGIKGHRGFSGLQGPPGSPGSPGEQGPSGASGPAGPRGPPGSAGSPGK 1141

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||:|| ||||||||||||||||
Db    1142 DGLNGLPGPIGPPGPRGRTGDSGPAGPPGPPGPPGPPGPP 1181


RESULT   4
I50629
collagen - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C;Accession: I50629
R;Fuller, F.; Boedtker, H.
Biochemistry 20, 996-1006, 1981
A;Title: Sequence determination and analysis of the 3' region of chicken pro-alpha 1(I)
A;Reference number: I50623; MUID:81160715
A;Accession: I50629
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-473 <FUL>
A;Cross-references: EMBL:V00401; NID:g63307; PIDN:CAA23695.1; PID:g63308
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F;244-473/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             96.2%;  Score 558;  DB 2;  Length 473;
  Best Local Similarity   95.0%;  Pred. No. 1.7e-29;
  Matches   95;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||||||||||||:||||||||||||||||||:||||||||||||||||||||||||| ||
Db     102 RGDKGETGEQGDRGMKGHRGFSGLQGPPGPPGAPGEQGPSGASGPAGPRGPPGSAGAAGK 161

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||: ||||||||||||||||||
Db     162 DGLNGLPGPIGPPGPRGRTGEVGPVGPPGPPGPPGPPGPP 201


RESULT   5
CGCH1S
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
C;Accession: A90458; A90181; A02857
R;Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross,
Biochemistry 21, 2048-2055, 1982
A;Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete pr
A;Reference number: A90458; MUID:82231995
A;Accession: A90458
A;Molecule type: protein
A;Residues: 1-1036 <HIG>
A;Experimental source: skin
A;Note: this is the latest in a series of papers from these workers elucidating the s
R;Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the
A;Reference number: A90181; MUID:72243016
A;Accession: A90181
A;Molecule type: protein
A;Residues: 1037-1042 <EYR>
A;Experimental source: skin
A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in s
C;Comment: Most of the prolines at the third position of the tripeptide repeating uni
C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trime
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental


  Query Match             96.2%;  Score 558;  DB 1;  Length 1042;
  Best Local Similarity   95.0%;  Pred. No. 3.1e-29;
  Matches   95;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||||||||||||:||||||||||||||||||:||||||||||||||||||||||||| ||
Db     931 RGDKGETGEQGDRGMKGHRGFSGLQGPPGPPGAPGEQGPSGASGPAGPRGPPGSAGAAGK 990

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||: ||||||||||||||||||
Db     991 DGLNGLPGPIGPPGPRGRTGEVGPVGPPGPPGPPGPPGPP 1030


RESULT   6
CGRT1S
collagen alpha 1(I) chain - rat (tentative sequence) (fragments)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C;Accession: A90559; A90552; A92029; A90353; A90566; A90357; A90362; A90379; A91209;
R;Bornstein, P.
Biochemistry 8, 63-71, 1969
A;Title: Comparative sequence studies of rat skin and tendon collagen. II. The absenc
A;Reference number: A90559; MUID:69155173
A;Contents: CNBr0 and CNBr1
A;Accession: A90559
A;Molecule type: protein
A;Residues: 1-19 <BO1>
A;Experimental source: tendon
```

A;Note: sequences from skin and tendon appear to be identical
A;Note: the amino-terminal tetrapeptide may be removed by limited proteolysis during ext
R;Kang, A.H.; Bornstein, P.; Piez, K.A.
Biochemistry 6, 788-795, 1967
A;Title: The amino acid sequence of peptides from the cross-linking region of rat skin c
A;Reference number: A90552; MUID:67162268
A;Contents: CNBr1
A;Accession: A90552
A;Molecule type: protein
A;Residues: 5-19 <KAN>
A;Experimental source: skin
R;Bornstein, P.
J. Biol. Chem. 242, 2572-2574, 1967
A;Title: The incomplete hydroxylation of individual prolyl residues in collagen.
A;Reference number: A92029; MUID:67165368
A;Contents: CNBr2
A;Accession: A92029
A;Molecule type: protein
A;Residues: 20-55 <BO2>
A;Experimental source: skin and tendon
R;Butler, W.T.; Ponds, S.L.
Biochemistry 10, 2076-2081, 1971
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90353; MUID:71263178
A;Contents: CNBr4
A;Accession: A90353
A;Molecule type: protein
A;Residues: 56-102 <BU1>
A;Experimental source: skin
R;Butler, W.T.
Biochemistry 9, 44-50, 1970
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The cov
A;Reference number: A90566; MUID:70085124
A;Contents: CNBr5
A;Accession: A90566
A;Molecule type: protein
A;Residues: 103-139 <BU2>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxyla
A;Reference number: A90357; MUID:72136131
A;Contents: CNBr8
A;Accession: A90357
A;Molecule type: protein
A;Residues: 140-238 <BA1>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxyla
A;Reference number: A90362; MUID:73006942
A;Contents: CNBr8
A;Accession: A90362
A;Molecule type: protein
A;Residues: 239-418 <BA2>
A;Experimental source: skin
R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2946-2953, 1974
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90379; MUID:74271984
A;Contents: CNBr3
A;Accession: A90379
A;Molecule type: protein
A;Residues: 419-567 <BU3>
A;Experimental source: skin
R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in neutr
A;Reference number: A91209; MUID:74011954
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <ST1>
A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chai
A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen)
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha1-chain.
A;Reference number: A91385; MUID:73049495
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <ST2>
A;Experimental source: skin
A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the seq
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1
C;Comment: Prolines and lysines at the third position of the tripeptide repeating uni
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin co
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydro
F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #st
F;9/Modified site: allysine (Lys) #status experimental
F;103,424,547/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (Lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 83.3%; Score 483; DB 1; Length 671;
Best Local Similarity 79.0%; Pred. No. 1.4e-24;
Matches 79; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

```
Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           :|| |  |  | :|  |  | |||:||||||||::||||||||||||||||||:|||
Db     547 KGDTGAPGAPGSQGAPGLZGMSGLZGPPGPPGSPGZZGPSGASGPAGPRGPPGSAGSPGK 606

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           :||:||||||||||||||||:||| ||||||||||||||
Db     607 BGLBGLPGPIGPPGPRGRTGBAGPSGPPGPPGPPGPPGPP 646
```

RESULT 7
A41182
collagen alpha 1(II) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C;Accession: A41182; A44885
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, a
A;Reference number: A41182; MUID:91358489
A;Accession: A41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1419 <MET>
A;Cross-references: GB:M65161
R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A;Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to carti
A;Reference number: A44885; MUID:91347939
A;Accession: A44885
A;Molecule type: DNA
A;Residues: 1-28 <CHE>
A;Cross-references: GB:S63190; NID:g234368; PIDN:AAB19627.1; PID:g234369
A;Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBIP:63192)
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; tr
F;1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 81.4%; Score 472; DB 2; Length 1419;
Best Local Similarity 78.0%; Pred. No. 1.3e-23;
Matches 78; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

```
Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          ||||||:||||:||:||||||:|||| ||||| |:|| || :||:|||||||  |  ||
Db   1047 RGDKGESGEQGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1106

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          || ||:|||||||||||||:|: ||||||| ||||||||||
Db   1107 DGSNGIPGPIGPPGPRGRSGETGPVGPPGSPGPPGPPGPP 1146


RESULT   8
B41182
collagen alpha 1(II) chain precursor (long splice form) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Jul-1999
C;Accession: B41182
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A;Reference number: A41182; MUID:91358489
A;Accession: B41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1487 <MET>
A;Cross-references: GB:M65161
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F;33-91/Domain: von Willebrand factor type C repeat homology <VWC>
F;1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             81.4%;  Score 472;  DB 2;  Length 1487;
  Best Local Similarity   78.0%;  Pred. No. 1.4e-23;
  Matches   78;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          ||||||:||||:||:||||||:|||| ||||| |:|| || :||:|||||||  |  ||
Db   1115 RGDKGESGEQGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1174

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          || ||:|||||||||||||:|: ||||||| ||||||||||
Db   1175 DGSNGIPGPIGPPGPRGRSGETGPVGPPGSPGPPGPPGPP 1214


RESULT   9
T45467
collagen alpha 1(II) chain precursor [imported] - horse
N;Alternate names: type II collagen
C;Species: Equus caballus (domestic horse)
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
C;Accession: T45467
R;Richardson, D.W.; Dodge, G.R.
submitted to the EMBL Data Library, June 1996
A;Description: Cloning of equine type II collagen and modulation of its expression in eq
A;Reference number: Z22977
A;Accession: T45467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1418 <RIC>
A;Cross-references: EMBL:U62528; PIDN:AAB05773.1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;


  Query Match             79.5%;  Score 461;  DB 2;  Length 1418;
  Best Local Similarity   76.0%;  Pred. No. 6.8e-23;
  Matches   76;  Conservative    9;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||| || |:||:||||||:|||| ||||| |:|| || :||:|||||||  |  ||
Db   1046 RGDKGEAGEAGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1105

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          || ||:|||||||||||||:|: || |||| ||||||||||
Db   1106 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP 1145


RESULT  10
CGHU6C
collagen alpha 1(II) chain precursor [validated] - human
N;Alternate names: procollagen alpha 1(II) chain
N;Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collag
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 08-Dec-2000
C;Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674;
7250; I37251; I37252; I37253; I37254; I55338; I59535; I61910
R;Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A;Title: The human type II procollagen gene: identification of an additional protein-
A;Reference number: A38513; MUID:91184811
A;Accession: A38513
A;Molecule type: DNA
A;Residues: 1-103 <RYA>
A;Cross-references: GB:M60299; NID:g180883; PIDN:AAA73873.1; PID:g180884
R;Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A;Title: Nucleotide sequence of the full length cDNA encoding for human type II proco
A;Reference number: S06715; MUID:90067946
A;Accession: S06715
A;Molecule type: mRNA
A;Residues: 1-28,'R',99-1487 <SU2>
A;Cross-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID:g29516
A;Note: alternative splice form 1
R;Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen,
Biochem. J. 285, 287-294, 1992
A;Title: Structural analysis of the regulatory elements of the type-II procollagen ge
A;Reference number: S24270; MUID:92344585
A;Accession: S24270
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-28 <VIK>
A;Cross-references: EMBL:X58709; GB:S40537; NID:g35659
A;Note: this translation is not annotated in GenBank entry HSPROCOE1, release 111.0
R;Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
Gene 44, 11-16, 1986
A;Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A;Reference number: A24828; MUID:87031574
A;Accession: A24828
A;Molecule type: DNA
A;Residues: 1-8,'T',10-28 <NUN>
A;Cross-references: GB:M25698; NID:g180872; PIDN:AAA52051.1; PID:g553237
R;Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A;Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(I
A;Reference number: S06496; MUID:90026318
A;Accession: S06496
A;Molecule type: mRNA
A;Residues: 7-28,'R',99-157,'P',159-440,'G',442-456,'E',458-640,'A',642-831,'PA',834,
A;Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
A;Note: alternative splice form 1
R;Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A;Title: Differential expression of a cysteine-rich domain in the amino-terminal prop
A;Reference number: A35428; MUID:90285153
A;Accession: A35428
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-81,'L',83-103 <RYA2>
A;Note: alternative splice form 2; splicing appears to be under developmental regulat
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A;Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide c
A;Reference number: A30147; MUID:89233138
A;Accession: A30147
A;Molecule type: DNA
A;Residues: 104-157,'P',159-236 <SUM>
A;Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168;
R;Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
```

Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A;Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
A;Reference number: A94227; MUID:90370826
A;Accession: A33116
A;Molecule type: DNA
A;Residues: 171-172,'C',174-175 <ALA>
A;Note: mutant sequence from a family with family with primary generalized osteoarthriti
R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
A;Reference number: S64673; MUID:96195147
A;Accession: S64674
A;Molecule type: protein
A;Residues: 188-189,'X',191-195;1224-1230,'X',1232-1236 <DIA>
R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
Eur. J. Biochem. 234, 125-131, 1995
A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil car
A;Reference number: S63514; MUID:96096730
A;Accession: S63514
A;Molecule type: protein
A;Residues: 243-261;575-590;756-763,'X',765-779 <FRA>
R;Tiller, G.E.; Weis, M.A.; Polumbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
Am. J. Hum. Genet. 56, 388-395, 1995
A;Title: An RNA-splicing mutation (G+5IVS20) in the type II collagen gene (COL2A1) in a
A;Reference number: I38867; MUID:95150028
A;Accession: I38867
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TIL1>
A;Cross-references: EMBL:U15195; NID:g557053; PIDN:AAB60370.1; PID:g557054
R;Ramirez, F.
submitted to the EMBL Data Library, December 1988
A;Reference number: S04892
A;Accession: S04892
A;Molecule type: mRNA
A;Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'F',836-1214 <RAM>
A;Cross-references: EMBL:X13783; NID:g30037; PIDN:CAA32030.1; PID:g930050
R;Vikkula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A;Title: Structural analyses of the polymorphic area in type II collagen gene.
A;Reference number: S05000; MUID:89325561
A;Accession: S05000
A;Molecule type: DNA
A;Residues: 630-640,'A',642-785 <VIK2>
A;Cross-references: EMBL:X16158; NID:g29951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
J. Biol. Chem. 267, 22522-22526, 1992
A;Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain prc
A;Reference number: A44309; MUID:93054548
A;Accession: A44309
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA; mRNA
A;Residues: 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',1038-1052,'E',1054-1068,'T',
A;Cross-references: GB:L00977; NID:g180812; PIDN:AAB23914.1; PID:g258774
A;Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence wer
A;Note: this translation is not annotated and this publication is not cited in GenBank e
A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individua
A;Reference number: S16502; MUID:90251662
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184,'GPSGKDGANGIPGPI',1185-1199 <TIL2>
A;Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>
A;Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II
A;Reference number: A27280; MUID:88067771
A;Accession: A27280
A;Molecule type: DNA; mRNA
A;Residues: 1175-1487 <ELI>
A;Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;1395-1408 <VAN>
A;Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal p
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human
A;Reference number: A21733; MUID:84118798
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295 <STR1>
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909,'PE' <STR2>
A;Cross-references: GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human
gene.
A;Reference number: A24561; MUID:86104139
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358 <NUN2>
A;Cross-references: GB:M12048; NID:g180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A;Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramir
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) c
A;Reference number: I37249; MUID:85215609
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28;'R',99-114;541-578;786-802;1055-1056,'N',1058-1068,'T',1070-1109;120
A;Accession: I84453
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 7-28 <SAN2>
A;Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I37250
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 541-560 <SAN3>
A;Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A;Accession: I37251

Query Match 79.5%; Score 461; DB 1; Length 1487;
Best Local Similarity 76.0%; Pred. No. 7e-23;
Matches 76; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

```
Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| || |:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db    1115 RGDKGEAGEPGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1174

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:|||||||||||||:|: || |||| ||||||||||
Db    1175 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPGP 1214
```

```
RESULT  11
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: B40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A;Reference number: A40333; MUID:92011898
A;Accession: B40333
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1486 <SUA>
A;Cross-references: GB:M63595
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VWC>
F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             77.8%;  Score 451;  DB 1;  Length 1486;
  Best Local Similarity   76.0%;  Pred. No. 3.1e-22;
  Matches   76;  Conservative    8;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| || |:|| ||||||:|||| |||||| |:|| :| :|||||||||  |  ||
Db    1117 RGDKGEAGEAGERGQKGHRGFTGLQGLPGPPGSAGDQGATGPAGPAGPRGPPGPVGPSGK 1176

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||: |||||||||||:|: || |||| ||||||||||
Db    1177 DGSNGISGPIGPPGPRGRSGETGPSGPPGQPGPPGPPGPP 1216


RESULT  12
A40333
collagen alpha 1'(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C;Accession: A40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A;Reference number: A40333; MUID:92011898
A;Accession: A40333
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1492 <SUA>
A;Cross-references: GB:M63596
A;Note: this sequence is presented as substitutions relative to another sequence in a fi
es they replace; the appropriate interpretation of the sequence figure was reconstructed
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VWC>
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             77.4%;  Score 449;  DB 2;  Length 1492;
  Best Local Similarity   74.0%;  Pred. No. 4.1e-22;
  Matches   74;  Conservative   11;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| || |:|| ||||||:|||| |||||: |:|| :|::||:|||||||  |  ||
Db    1120 RGDKGEAGEAGERGQKGHRGFTGLQGLPGPPGTAGDQGATGSAGPSGPRGPPGPVGPSGK 1179

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:||||||||||||:|: || ||||  |||||||||
Db    1180 DGSNGIPGPIGPPGPRGRSGETGPAGPPGQSGPPGPPGPP 1219


RESULT  13
S59513
collagen II A1 protein - zebra fish (fragment)
C;Species: Brachydanio rerio (zebra fish)
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 13-Aug-1999
C;Accession: S59513
R;Yan, Y.; Hotta, K.; Riggleman, B.; Postlethwait, J.H.
submitted to the EMBL Data Library, March 1995
A;Description: Expression of a type II collagen gene in the zebrafish embryonic axis.
A;Reference number: S59513
A;Accession: S59513
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-464 <YAN>
A;Cross-references: EMBL:U23822; NID:g773660; PIDN:AAA96815.1; PID:g773661
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
F;236-464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             77.1%;  Score 447;  DB 2;  Length 464;
  Best Local Similarity   75.0%;  Pred. No. 2.2e-22;
  Matches   75;  Conservative   10;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| || |:|| ||||||:|||| |||||||:|| :| :||:| :|| |  |  ||
Db      92 RGDKGEAGEAGERGQKGHRGFTGLQGLPGPPGSPGDQGAAGPAGPSGAKGPSGPVGPAGK 151

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || || ||||||||||||:|::|||||| ||||||||||
Db     152 DGSNGQPGPIGPPGPRGRSGESGPVGPPGNPGPPGPPGPP 191


RESULT  14
S07133
collagen alpha 1(II) chain precursor - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Jan-1995 #sequence_revision 17-Jul-1998 #text_change 13-Aug-1999
C;Accession: S07133; S24447
R;Deak, F.; Argraves, W.S.; Kiss, I.; Sparks, K.J.; Goetinck, P.F.
Biochem. J. 229, 189-196, 1985
A;Title: Primary structure of the telopeptide and a portion of the helical domain of
A;Reference number: S07133; MUID:85306862
A;Accession: S07133
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-193 <DEA>
A;Cross-references: EMBL:X02663; NID:g63318; PIDN:CAA26499.1; PID:g755732
R;van der Rest, M.; Mayne, R.
J. Biol. Chem. 263, 1615-1618, 1988
A;Title: Type IX collagen proteoglycan from cartilage is covalently cross-linked to t
A;Reference number: S23813; MUID:88115274
A;Accession: S24447
A;Molecule type: protein
A;Residues: 124-125,'X',127-129,'AQ' <RES>
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo


  Query Match             75.7%;  Score 439;  DB 2;  Length 193;
  Best Local Similarity   75.0%;  Pred. No. 3.5e-22;
  Matches   72;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       5 GETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGKDGLN 64
           ||||| |:||:||||||:|||| |||||  |:|| :| :||:|||||||  |  |||| |
Db       1 GETGEAGERGLKGHRGFTGLQGLPGPPGPSGDQGAAGPAGPSGPRGPPGPVGPSGKDGSN 60

Qy      65 GLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           |:||||||||||||:|: || |||| ||||||||||
Db      61 GMPGPIGPPGPRGRSGEPGPAGPPGNPGPPGPPGPP 96


RESULT  15
I49607
procollagen type V alpha 2 - mouse
C;Species: Mus musculus (house mouse)
```

```
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C;Accession: I49607
R;Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A;Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
A;Reference number: I49607; MUID:93214071
A;Accession: I49607
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1497 <RES>
A;Cross-references: GB:L02918; NID:g309180; PIDN:AAA37440.1; PID:g309181
C;Genetics:
A;Gene: Col5a-2
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F;39-98/Domain: von Willebrand factor type C repeat homology <VWC>
F;1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>


  Query Match             70.7%;  Score 410;  DB 2;  Length 1497;
  Best Local Similarity   71.0%;  Pred. No. 1.3e-19;
  Matches   71;  Conservative    8;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||: |::|||| |||||:|||| |||||  |||| :|  || |||||||  |: ||
Db    1125 RGDKGDNGDRGDRGQKGHRGFTGLQGLPGPPGPNGEQGSAGIPGPFGPRGPPGPVGSSGK 1184

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           :|  |  ||||||| ||  |:||| |||| ||||||||||
Db    1185 EGNPGPLGPIGPPGVRGSVGEAGPEGPPGEPGPPGPPGPP 1224


Search completed: January 28, 2002, 07:47:27
Job time: 32 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:48:32 ; Search time 15.36 Seconds
(without alignments)
238.703 Million cell updates/sec

Title: US-09-710-239-29
Perfect score: 580
Sequence: 1 RGDKGETGEQGDRGIKGHRG.....DAGPVGPPGPPGPPGPPGPP 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	580	100.0	1464	1	CA11_HUMAN	P02452 homo sapien
2	577	99.5	1460	1	CA11_CANFA	Q9xsj7 canis famil
3	565	97.4	779	1	CA11_BOVIN	P02453 bos taurus
4	562	96.9	1453	1	CA11_MOUSE	P11087 mus musculu
5	558	96.2	1453	1	CA11_CHICK	P02457 gallus gall
6	484	83.4	671	1	CA11_RAT	P02454 rattus norv
7	472	81.4	1459	1	CA12_MOUSE	P28481 mus musculu
8	461	79.5	1418	1	CA12_HUMAN	P02458 homo sapien
9	439	75.7	369	1	CA12_CHICK	P02460 gallus gall
10	404	69.7	1496	1	CA25_HUMAN	P05997 homo sapien
11	400	69.0	1464	1	CA13_MOUSE	P08121 mus musculu
12	399	68.8	636	1	CA13_RAT	P13941 rattus norv
13	395	68.1	1362	1	CA21_CHICK	P02467 gallus gall
14	392	67.6	1466	1	CA13_HUMAN	P02461 homo sapien
15	390	67.2	1049	1	CA13_BOVIN	P04258 bos taurus
16	373	64.3	1355	1	CA21_RANCA	O42350 rana catesb
17	359	61.9	1372	1	CA21_MOUSE	Q01149 mus musculu
18	357	61.6	1366	1	CA21_CANFA	O46392 canis famil
19	353	60.9	1364	1	CA21_BOVIN	P02465 bos taurus
20	349	60.2	526	1	CA21_RABIT	Q28668 oryctolagus
21	349	60.2	1372	1	CA21_RAT	P02466 rattus norv
22	348	60.0	1366	1	CA21_HUMAN	P08123 homo sapien
23	337	58.1	1356	1	CA21_ONCMY	O93484 oncorhynchu
24	336	57.9	1262	1	CA13_CHICK	P12105 gallus gall
25	313.5	54.1	1650	1	CA2B_MOUSE	Q64739 mus musculu
26	313.5	54.1	1736	1	CA2B_HUMAN	P13942 homo sapien
27	307.5	53.0	1838	1	CA15_HUMAN	P20908 homo sapien
28	299.5	51.6	1804	1	CA1B_MOUSE	Q61245 mus musculu
29	298	51.4	675	1	CA39_CHICK	P32017 gallus gall
30	298	51.4	1527	1	CA1H_MOUSE	P39061 mus musculu
31	297	51.2	1669	1	CA14_HUMAN	P02462 homo sapien
32	296.5	51.1	1806	1	CA1B_HUMAN	P12107 homo sapien
33	295.5	50.9	1669	1	CA14_MOUSE	P02463 mus musculu
34	295	50.9	911	1	CA1B_BOVIN	Q28083 bos taurus
35	293.5	50.6	366	1	CAS4_EPHMU	P18503 ephydatia m
36	293	50.5	1027	1	CAFF_RIFPA	P30754 riftia pach
37	290	50.0	3124	1	CA1C_CHICK	P13944 gallus gall
38	289	49.8	623	1	CA44_RABIT	P55787 oryctolagus
39	287.5	49.6	674	1	CA1A_CHICK	P08125 gallus gall
40	286.5	49.4	1685	1	CA54_HUMAN	P29400 homo sapien
41	286.5	49.4	1690	1	CA44_HUMAN	P53420 homo sapien
42	286.5	49.4	2944	1	CA17_HUMAN	Q02388 homo sapien
43	285.5	49.2	674	1	CA1A_BOVIN	P23206 bos taurus
44	283.5	48.9	1516	1	CA1H_HUMAN	P39060 homo sapien
45	283	48.8	482	1	CA1B_RAT	P20909 rattus norv

ALIGNMENTS

```
RESULT   1
CA11_HUMAN
ID   CA11_HUMAN     STANDARD;      PRT;  1464 AA.
AC   P02452; Q15176; Q14037;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-472 FROM N.A.
RX   MEDLINE=89025644; PubMed=3178743;
RA   Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA   Jaenisch R., Prockup D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT   chain of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [2]
RP   SEQUENCE OF 1-181 FROM N.A.
RX   MEDLINE=84270697; PubMed=6462220;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA   Myers J., Williams C., Ramirez F.;
RT   "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT   conservation of a pattern of introns and exons.";
RL   Nature 310:337-340(1984).
RN   [3]
RP   SEQUENCE OF 162-301.
RC   TISSUE=Skin;
RX   MEDLINE=71038625; PubMed=5529814;
RA   Click E.M., Bornstein P.;
RT   "Isolation and characterization of the cyanogen bromide peptides from
RT   the alpha 1 and alpha 2 chains of human skin collagen.";
RL   Biochemistry 9:4699-4706(1970).
RN   [4]
RP   SEQUENCE OF 263-268.
RC   TISSUE=Skin;
RX   MEDLINE=71001508; PubMed=4319110;
RA   Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT   "A comparative study of glycopeptides derived from selected
RT   vertebrate collagens. A possible role of the carbohydrate in fibril
RT   formation.";
RL   J. Biol. Chem. 245:5042-5048(1970).
RN   [5]
RP   SEQUENCE OF 425-1464 FROM N.A.
RX   MEDLINE=84080385; PubMed=6689127;
RA   Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA   Prockop D.J.;
RT   "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT   pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT   of structures that are conserved during evolution.";
RL   Biochemistry 22:5213-5223(1983).
RN   [6]
```

RP SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=88124208; PubMed=3340531;
RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [7]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to
RT transcriptional control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [8]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [9]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.J.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
RN [13]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Dalgleish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
RN [14]
RP VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT change in one human pro alpha 1(I) collagen allele.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN [15]
RP VARIANT OI-II ARG-569.
RX MEDLINE=87222295; PubMed=3108247;
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT collagen.";
RL J. Biol. Chem. 262:7021-7027(1987).
RN [16]
RP VARIANT OI-II CYS-926.
RX MEDLINE=88033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT "A point mutation in a type I procollagen gene converts glycine 748
RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT a lethal variant of osteogenesis imperfecta.";
RL J. Biol. Chem. 262:14737-14744(1987).
RN [17]
RP VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT the peptide defect by in vitro expression of the mutant cDNA.";
RL J. Biol. Chem. 263:11627-11630(1988).
RN [18]
RP VARIANT OI CYS-1195.
RX MEDLINE=89218628; PubMed=3244312;
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT "A cysteine for glycine substitution at position 1017 in an alpha
RT 1(I) chain of type I collagen in a patient with mild dominantly
RT inherited osteogenesis imperfecta.";
RL Mol. Biol. Med. 5:197-207(1988).
RN [19]
RP VARIANT OI-II VAL-434.
RX MEDLINE=89255493; PubMed=2470760;
RA Patterson E., Smiley E., Bonadio J.;
RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT mutation.";
RL J. Biol. Chem. 264:10083-10087(1989).
RN [20]
RP VARIANT OI-IV SER-1010.
RX MEDLINE=89308591; PubMed=2745420;
RA Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL J. Biol. Chem. 264:11893-11900(1989).
RN [21]
RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX MEDLINE=89380165; PubMed=2777764;
RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT "Characterization of point mutations in the collagen COL1A1 and
RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL J. Biol. Chem. 264:15809-15812(1989).
RN [22]
RP VARIANT OI SER-1022.
RX MEDLINE=90062068; PubMed=2511192;
RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT variant of osteogenesis imperfecta minimally destabilizes the triple
RT helix of type I procollagen. The effects of glycine substitutions on
RT thermal stability are either position of amino acid specific.";
RL J. Biol. Chem. 264:19694-19699(1989).
RN [23]
RP VARIANT OI-II CYS-1082.
RX MEDLINE=89109573; PubMed=2913053;
RA Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT "A lethal variant of osteogenesis imperfecta has a single base
RT mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT chain of type I procollagen. The asymptomatic mother has an
RT unidentified mutation producing an overmodified and unstable type I
RT procollagen.";
RL J. Clin. Invest. 83:574-584(1989).
RN [24]
RP VARIANT OI CYS-272; CYS-704 AND CYS-896.
RX MEDLINE=90009313; PubMed=2794057;
RA Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
RA Weiss L., Graham J.M., Byers P.H.;
RT "Osteogenesis imperfecta. The position of substitution for glycine by
RT cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT type I collagen determines the clinical phenotype.";
RL J. Clin. Invest. 84:1206-1214(1989).
RN [25]
RP VARIANT OI-II CYS-422.

```
  Query Match           100.0%;  Score 580;  DB 1;  Length 1464;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-27;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1093 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 1152

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||||||||||||||||||||||||||||||||||||||||||
Db    1153 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 1192


RESULT   2
CA11_CANFA
ID   CA11_CANFA     STANDARD;      PRT;  1460 AA.
AC   Q9XSJ7;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
RT   "Sequence of normal canine COL1A1 cDNA.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF153062; AAD34619.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   PROPEP       23    157       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN       158   1214       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP     1215   1460       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       34     92       VWFC.
FT   DOMAIN      158    174       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN      175   1188       TRIPLE-HELICAL REGION.
FT   DOMAIN     1189   1214       NONHELICAL REGION (C-TERMINAL).
FT   SITE        741    743       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE       1089   1091       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD   1361   1361       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1460 AA;  138762 MW;  58E3674D2B570697 CRC64;

  Query Match           99.5%;  Score 577;  DB 1;  Length 1460;
  Best Local Similarity  99.0%;  Pred. No. 2.6e-27;
  Matches   99;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db    1089 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGSPGK 1148

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           |||||||||||||||||||||||||||||||||||||||||
Db    1149 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 1188


RESULT   3
CA11_BOVIN
ID   CA11_BOVIN     STANDARD;      PRT;   779 AA.
AC   P02453;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN (FRAGMENTS).
GN   COL1A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-19.
RX   MEDLINE=72255334; PubMed=4115172;
RA   Rauterberg J., Timpl R., Furthmayr H.;
RT   "Structural characterization of N-terminal antigenic determinants in
RT   calf and human collagen.";
RL   Eur. J. Biochem. 27:231-237(1972).
RN   [2]
RP   SEQUENCE OF 20-145.
RX   MEDLINE=76022320; PubMed=1164916;
RA   Fietzek P.P., Kuehn K.;
RT   "The covalent structure of collagen: amino-acid sequence of the
RT   cyanogen-bromide peptides alpha-1-CB2, alpha-1-CB4 and alpha-1-CB5
RT   from calf-skin collagen.";
RL   Eur. J. Biochem. 52:77-82(1975).
RN   [3]
RP   SEQUENCE OF 146-294.
RX   MEDLINE=73049499; PubMed=4673951;
RA   Fietzek P.P., Wendt P., Kell I., Kuehn K.;
RT   "The covalent structure of collagen: amino acid sequence of alpha-1-
RT   CB3 from calf skin collagen.";
RL   FEBS Lett. 26:74-76(1972).
RN   [4]
RP   SEQUENCE OF 295-562.
RX   MEDLINE=74086118; PubMed=4359390;
RA   Fietzek P.P., Rexrodt F.W., Hopper K.E., Kuehn K.;
RT   "The covalent structure of collagen. 2. The amino-acid sequence of
RT   alpha-1-CB7 from calf-skin collagen.";
RL   Eur. J. Biochem. 38:396-400(1973).
RN   [5]
RP   SEQUENCE OF 563-675.
RX   MEDLINE=73042276; PubMed=4343808;
RA   Wendt P., Mark K.V.D., Rexrodt F., Kuehn K.;
RT   "The covalent structure of collagen. The amino-acid sequence of the
RT   112-residues. Amino-terminal part of peptide alpha-1-CB6 from calf-
RT   skin collagen.";
RL   Eur. J. Biochem. 30:169-183(1972).
RN   [6]
RP   SEQUENCE OF 676-751.
RX   MEDLINE=73042275; PubMed=4343807;
RA   Fietzek P.P., Rexrodt F.W., Wendt P., Stark M., Kuehn K.;
RT   "The covalent structure of collagen. Amino-acid sequence of peptide
RT   alpha-1-CB6-C2.";
RL   Eur. J. Biochem. 30:163-168(1972).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
```

```
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC       O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE.
CC   -!- MISCELLANEOUS: THE COMPLETE CHAIN CONTAINS 1052 RESIDUES.
DR   PIR; A91193; CGBO1S.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 12.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       9      9       CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                                INVOLVED IN CROSS-LINKING.
FT   MOD_RES     103    103       HYDROXYLATION.
FT   CARBOHYD    103    103       O-LINKED (GAL. . .).
FT   MOD_RES     115    115       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     124    124       HYDROXYLATION (POTENTIAL).
FT   NON_CONS    145    146
FT   MOD_RES     274    274       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     346    346       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     424    424       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     496    496       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     658    658       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     670    670       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     726    726       HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE
FT                                ONLY HYDROXYLATED PRO IN POSITION X (IN
FT                                THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
SQ   SEQUENCE   779 AA;  70346 MW;  E554A7FF084283D1 CRC64;
```

```
  Query Match             97.4%;  Score 565;  DB 1;  Length 779;
  Best Local Similarity   93.0%;  Pred. No. 8e-27;
  Matches   93;  Conservative    6;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||:||:||::|:||||||||||||||||||||||||||||||||||||||||||:|||
Db     655 RGBKGZTGZZGBRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGSPGK 714

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           |||||||||||||||||||||| ||||||||||||||||
Db     715 DGLNGLPGPIGPPGPRGRTGDAGPAGPPGPPGPPGPPGPP 754
```

```
RESULT   4
CA11_MOUSE
ID   CA11_MOUSE     STANDARD;      PRT;  1453 AA.
AC   P11087; Q60635;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1 OR COLA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N;
RX   MEDLINE=96033240; PubMed=8535610;
RA   Li S.W., Khillan J., Prockop D.J.;
RT   "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT   of type I procollagen.";
RL   Matrix Biol. 14:593-595(1995).
RN   [2]
RP   SEQUENCE OF 518-1128 FROM N.A.
RX   MEDLINE=86137403; PubMed=3841523;
RA   French B.T., Lee W.-H., Maul G.G.;
RT   "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT   collagen protein.";
RL   Gene 39:311-312(1985).
RN   [3]
RP   SEQUENCE OF 735-1130 FROM N.A.
RX   MEDLINE=83141374; PubMed=6298597;
RA   Monson J.M., Friedman J., McCarthy B.J.;
RT   "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
RT   evidence for a mouse B1 element within the gene.";
RL   Mol. Cell. Biol. 2:1362-1371(1982).
RN   [4]
RP   SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX   MEDLINE=83157109; PubMed=6219867;
RA   Monson J.M., McCarthy B.J.;
RT   "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT   evidence for insertions or deletions in gene coding sequences.";
RL   DNA 1:59-69(1981).
RN   [5]
RP   SEQUENCE OF 1442-1453 FROM N.A.
RX   MEDLINE=88124276; PubMed=3340560;
RA   Mooslehner K., Harbers K.;
RT   "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT   of the 3'-untranslated region.";
RL   Nucleic Acids Res. 16:773-773(1988).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08020; AAA88912.1; -.
DR   EMBL; X15896; CAA33904.1; -.
DR   EMBL; M14423; AAA37333.1; -.
DR   EMBL; M17491; AAA37334.1; -.
DR   EMBL; X06753; CAA29927.1; -.
DR   EMBL; K03036; AAA37332.1; -.
DR   EMBL; K03029; AAA37332.1; JOINED.
DR   EMBL; K03030; AAA37332.1; JOINED.
DR   EMBL; K03031; AAA37332.1; JOINED.
DR   EMBL; K03032; AAA37332.1; JOINED.
DR   EMBL; K03033; AAA37332.1; JOINED.
DR   EMBL; K03034; AAA37332.1; JOINED.
DR   EMBL; K03035; AAA37332.1; JOINED.
DR   PIR; A23982; A23982.
DR   MGD; MGI:88467; Cola1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     22
FT   PROPEP       23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN       152   1207       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP     1208   1453       CARBOXYL-TERMINAL PROPEPTIDE.
```

```
FT   DOMAIN       29     87       VWFC.
FT   DOMAIN      152    167       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN      168   1181       TRIPLE-HELICAL REGION.
FT   DOMAIN     1182   1207       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1354   1354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        734    736       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE       1082   1084       CELL ATTACHMENT SITE (POTENTIAL).
FT   CONFLICT   1450   1450       A -> V (IN REF. 5).
SQ   SEQUENCE   1453 AA;  137944 MW;  3B802E535DF81808 CRC64;


  Query Match             96.9%;  Score 562;  DB 1;  Length 1453;
  Best Local Similarity   96.0%;  Pred. No. 1.9e-26;
  Matches   96;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||:|||
Db    1082 RGDKGETGEQGDRGIKGHRGFSGLQGPPGSPGSPGEQGPSGASGPAGPRGPPGSAGSPGK 1141

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           |||||||||||||||||||:|| ||||||||||||||
Db    1142 DGLNGLPGPIGPPGPRGRTGDSGPAGPPGPPGPPGPPGPP 1181


RESULT   5
CA11_CHICK
ID   CA11_CHICK     STANDARD;      PRT;  1453 AA.
AC   P02457;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-153 FROM N.A.
RX   MEDLINE=88056316; PubMed=3678834;
RA   Finer M.H., Boedtker H., Doty P.;
RT   "Construction and characterization of cDNA clones encoding the 5' end
RT   of the chicken pro alpha 1(I) collagen mRNA.";
RL   Gene 56:71-78(1987).
RN   [2]
RP   SEQUENCE OF 1-144 FROM N.A.
RX   MEDLINE=88007542; PubMed=2820966;
RA   Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT   "Unusual DNA sequences located within the promoter region and the
RT   first intron of the chicken pro-alpha 1(I) collagen gene.";
RL   J. Biol. Chem. 262:13323-13332(1987).
RN   [3]
RP   SEQUENCE OF 152-1187.
RX   MEDLINE=82231995; PubMed=7093229;
RA   Highberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA   Kang A.H., Gross J.;
RT   "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the
RT   complete primary structure of the helical portion of the chick skin
RT   collagen alpha 1(I) chain.";
RL   Biochemistry 21:2048-2055(1982).
RN   [4]
RP   SEQUENCE OF 1200-1205.
RX   MEDLINE=72243016; PubMed=5047697;
RA   Eyre D.R., Glimcher M.J.;
RT   "Evidence for a previously undetected sequence at the carboxyterminus
RT   of the alpha 1 chain of chicken bone collagen.";
RL   Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN   [5]
RP   SEQUENCE OF 981-1453 FROM N.A.
RX   MEDLINE=81160715; PubMed=6927845;
RA   Fuller F., Boedtker H.;
RT   "Sequence determination and analysis of the 3' region of chicken pro-
RT   alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT   including the carboxy-terminal propeptide sequences.";
RL   Biochemistry 20:996-1006(1981).
RN   [6]
RP   SEQUENCE OF 1311-1453 FROM N.A.
RX   MEDLINE=80134546; PubMed=6987088;
RA   Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA   Pastan I., Decrombrugghe B., Fietzek P.P., Olsen B.R.;
RT   "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT   carboxyl end of pro alpha 1(I)-chains.";
RL   FEBS Lett. 111:61-65(1980).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17839; AAA48704.1; -.
DR   EMBL; M17838; AAA48704.1; JOINED.
DR   EMBL; V00401; CAA23695.1; -.
DR   EMBL; M10571; AAA48671.1; ALT_SEQ.
DR   EMBL; M17607; AAA48672.1; -.
DR   PIR; A02857; CGCH1S.
DR   PIR; A27179; A27179.
DR   PIR; A29367; A29367.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL       1     22
FT   PROPEP      23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN      152   1205       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP     1206   1453       C-TERMINAL PROPEPTIDE.
FT   DOMAIN       31     89       VWFC.
FT   MOD_RES     152    152       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES     254    254       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     851    851       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    1081   1081       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    1097   1097       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    1153   1153       HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE
FT                                ONLY HYDROXYLATED PRO IN POSITION X (IN
FT                                THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
FT   CONFLICT   1187   1187       F -> L (IN REF. 5).
FT   CONFLICT   1441   1441       Q -> H (IN REF. 6).
SQ   SEQUENCE   1453 AA;  137789 MW;  3BC6152134271F4D CRC64;


  Query Match             96.2%;  Score 558;  DB 1;  Length 1453;
  Best Local Similarity   95.0%;  Pred. No. 3.2e-26;
  Matches   95;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
```

```
          ||||||||||||||:||||||||||||||||||:||||||||||||||||||||||| ||
Db   1082 RGDKGETGEQGDRGMKGHRGFSGLQGPPGPPGAPGEQGPSGASGPAGPRGPPGSAGAAGK 1141

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPGPP 100
          ||||||||||||||||||: ||||||||||||||||||
Db   1142 DGLNGLPGPIGPPGPRGRTGEVGPVGPPGPPGPPGPPGPPGPP 1181


RESULT   6
CA11_RAT
ID   CA11_RAT       STANDARD;      PRT;   671 AA.
AC   P02454; P02455;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN (FRAGMENTS).
GN   COL1A1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-19.
RX   MEDLINE=69155173; PubMed=5777344;
RA   Bornstein P.;
RT   "Comparative sequence studies of rat skin and tendon collagen. II.
RT   The absence of a short sequence at the amino terminus of the skin
RT   alpha-1 chain.";
RL   Biochemistry 8:63-71(1969).
RN   [2]
RP   SEQUENCE OF 5-19.
RX   MEDLINE=67162268; PubMed=5337886;
RA   Kang A.H., Bornstein P., Piez K.A.;
RT   "The amino acid sequence of peptides from the cross-linking region of
RT   rat skin collagen.";
RL   Biochemistry 6:788-795(1967).
RN   [3]
RP   SEQUENCE OF 20-55.
RX   MEDLINE=67165368; PubMed=4290711;
RA   Bornstein P.;
RT   "The incomplete hydroxylation of individual prolyl residues in
RT   collagen.";
RL   J. Biol. Chem. 242:2572-2574(1967).
RN   [4]
RP   SEQUENCE OF 56-102.
RX   MEDLINE=71263178; PubMed=4327399;
RA   Butler W.T., Ponds S.L.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. Amino acid sequence of alpha 1-CB4.";
RL   Biochemistry 10:2076-2081(1971).
RN   [5]
RP   SEQUENCE OF 103-139.
RX   MEDLINE=70085124; PubMed=5411206;
RA   Butler W.T.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. The covalent structure of alpha 1-CB5, the major
RT   hexose-containing cyanogen bromide peptide of alpha 1.";
RL   Biochemistry 9:44-50(1970).
RN   [6]
RP   SEQUENCE OF 140-238.
RX   MEDLINE=72136131; PubMed=4335087;
RA   Balian G., Click E.M., Bornstein P.;
RT   "Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of
RT   the hydroxylamine-produced fragment HA1.";
RL   Biochemistry 10:4470-4478(1971).
RN   [7]
RP   SEQUENCE OF 239-418.
RX   MEDLINE=73006942; PubMed=4342027;
RA   Balian G., Click E.M., Hermodson M.A., Bornstein P.;
RT   "Structure of rat skin collagen alpha 1-CBB. Amino acid sequence of
RT   the hydroxyl amine-produced fragment HA2.";
RL   Biochemistry 11:3798-3806(1972).
RN   [8]
RP   SEQUENCE OF 419-567.
RX   MEDLINE=74271984; PubMed=4366532;
RA   Butler W.T., Underwood S.P., Finch J.E. Jr.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. Amino acid sequence of alpha 1-CB3.";
RL   Biochemistry 13:2946-2953(1974).
RN   [9]
RP   SEQUENCE OF 568-651.
RX   MEDLINE=74011954; PubMed=4126850;
RA   Stoltz M., Timpl R., Furthmayr H., Kuehn K.;
RT   "Structural and immunogenic properties of a major antigenic
RT   determinant in neutral salt-extracted rat-skin collagen.";
RL   Eur. J. Biochem. 37:287-294(1973).
RN   [10]
RP   SEQUENCE OF 651-671.
RX   MEDLINE=73049495; PubMed=4636751;
RA   Stoltz M., Timpl R., Kuehn K.;
RT   "Non-helical regions in rat collagen alpha 1-chain.";
RL   FEBS Lett. 26:61-65(1972).
RN   [11]
RP   SEQUENCE OF 529-567 FROM N.A.
RX   MEDLINE=85122694; PubMed=6395893;
RA   Genovese C., Rowe D., Kream B.;
RT   "Construction of DNA sequences complementary to rat alpha 1 and alpha
RT   2 collagen mRNA and their use in studying the regulation of type I
RT   collagen synthesis by 1,25-dihydroxyvitamin D.";
RL   Biochemistry 23:6210-6216(1984).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC       O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11432; AAA40832.1; ALT_SEQ.
DR   PIR; A02854; CGRT1S.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 10.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT   MOD_RES       9      9       CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                                INVOLVED IN CROSS-LINKING.
FT   MOD_RES      28     28       HYDROXYLATION (PROBABLE).
FT   MOD_RES      31     31       HYDROXYLATION (PROBABLE).
FT   MOD_RES      34     34       HYDROXYLATION (PROBABLE).
FT   MOD_RES      43     43       HYDROXYLATION (PROBABLE).
FT   MOD_RES      46     46       HYDROXYLATION (PROBABLE).
FT   MOD_RES      49     49       HYDROXYLATION (PROBABLE).
FT   MOD_RES     103    103       HYDROXYLATION.
FT   CARBOHYD    103    103       O-LINKED (GAL. . .).
FT   MOD_RES     424    424       HYDROXYLATION (PROBABLE).
FT   MOD_RES     547    547       HYDROXYLATION (PROBABLE).
FT   NON_CONS    567    568
FT   DOMAIN      641    651       MAJOR ANTIGENIC DETERMINANT (OF NEUTRAL
FT                                SALT-EXTRACTED RAT SKIN COLLAGEN).
SQ   SEQUENCE   671 AA;  60615 MW;  9DC3114204AC4918 CRC64;


  Query Match             83.4%;  Score 484;  DB 1;  Length 671;
```

```
  Best Local Similarity   79.0%;  Pred. No. 3.6e-22;
  Matches   79;  Conservative   10;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           :|| |  |  | :|  | :| |||:||||||||::|||||||||||||||||||:||| 
Db     547 KGDTGAPGAPGSQGAPGLQGMSGLZGPPGPPGSPGZZGPSGASGPAGPRGPPGSAGSPGK 606

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           :||:||||||||||||||||:||| ||||||||||||||
Db     607 BGLBGLPGPIGPPGPRGRTGBAGPSGPPGPPGPPGPPGPP 646


RESULT   7
CA12_MOUSE
ID   CA12_MOUSE     STANDARD;      PRT;  1459 AA.
AC   P28481;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].
GN   COL2A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91358489; PubMed=1885613;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. Biol. Chem. 266:16862-16869(1991).
RN   [2]
RP   SEQUENCE OF 1455-1459 FROM N.A.
RX   MEDLINE=91274355; PubMed=2054384;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Specific hybridization probes for mouse type I, II, III and IX
RT   collagen mRNAs.";
RL   Biochim. Biophys. Acta 1089:241-243(1991).
CC   -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M65161; AAA68100.1; -.
DR   EMBL; X57982; CAA41047.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   PROPEP       26    153       AMINO-TERMINAL PROPEPTIDE
FT                                (BY SIMILARITY).
FT   CHAIN       154   1213       COLLAGEN ALPHA 1(II) CHAIN.
FT   PROPEP     1214   1459       CARBOXYL-TERMINAL PROPEPTIDE
FT                                (CHONDROCALCIN).
FT   DOMAIN       32     89       VWFC.
FT   DOMAIN      173   1186       TRIPLE-HELICAL REGION.
FT   DOMAIN     1187   1213       NONHELICAL REGION (C-TERMINAL).
FT   VARSPLIC     29     29       Q -> R (IN SHORT ISOFORM).
FT   VARSPLIC     30     98       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   1459 AA;  139154 MW;  F6C84FA7C532E7F2 CRC64;


  Query Match             81.4%;  Score 472;  DB 1;  Length 1459;
  Best Local Similarity   78.0%;  Pred. No. 3.2e-21;
  Matches   78;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||:||||:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db    1087 RGDKGESGEQGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1146

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:|||||||||||||:|: ||||||| ||||||||||
Db    1147 DGSNGIPGPIGPPGPRGRSGETGPVGPPGSPGPPGPPGPP 1186


RESULT   8
CA12_HUMAN
ID   CA12_HUMAN     STANDARD;      PRT;  1418 AA.
AC   P02458;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90067946; PubMed=2587267;
RA   Su M.W., Lee B., Ramirez F., Machado M., Horton W.;
RT   "Nucleotide sequence of the full length cDNA encoding for human type
RT   II procollagen.";
RL   Nucleic Acids Res. 17:9473-9473(1989).
RN   [2]
RP   SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE=87031574; PubMed=3021582;
RA   Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
RT   "Promoter region of the human pro-alpha 1(II)-collagen gene.";
RL   Gene 44:11-16(1986).
RN   [3]
RP   SEQUENCE OF 432-1145 FROM N.A.
RA   Ramirez F.;
RL   Submitted (DEC-1988) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 963-1418 FROM N.A.
RX   MEDLINE=85190534; PubMed=3857598;
RA   Cheah K.S.E., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
RT   "Identification and characterization of the human type II collagen
RT   gene (COL2A1).";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN   [5]
RP   SEQUENCE OF 1120-1398 FROM N.A.
RX   MEDLINE=85306861; PubMed=3840017;
RA   Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,
RA   Vuorio E.;
RT   "Construction and identification of a cDNA clone for human type II
RT   procollagen mRNA.";
RL   Biochem. J. 229:183-188(1985).
RN   [6]
RP   SEQUENCE OF 1106-1418 FROM N.A.
RX   MEDLINE=88067771; PubMed=2825137;
RA   Elima K., Vuorio T., Vuorio E.;
RT   "Determination of the single polyadenylation site of the human pro
RT   alpha 1(II) collagen gene.";
```

RL Nucleic Acids Res. 15:9499-9504(1987).
RN [7]
RP SEQUENCE OF 1227-1289 FROM N.A.
RX MEDLINE=86104139; PubMed=3002437;
RA Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
RT "Isolation and partial characterization of genomic clones coding for
RT a human pro-alpha 1 (II) collagen chain and demonstration of
RT restriction fragment length polymorphism at the 3' end of the gene.";
RL Biochemistry 24:6343-6348(1985).
RN [8]
RP SEQUENCE OF 1176-1226 FROM N.A.
RX MEDLINE=84118798; PubMed=6320112;
RA Strom C.M., Upholt W.B.;
RT "Isolation and characterization of genomic clones corresponding to
RT the human type II procollagen gene.";
RL Nucleic Acids Res. 12:1025-1038(1984).
RN [9]
RP SEQUENCE OF 35-167 FROM N.A.
RX MEDLINE=89233138; PubMed=2714801;
RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
RT "Organization of the exons coding for pro alpha 1(II) collagen N-
RT propeptide confirms a distinct evolutionary history of this domain of
RT the fibrillar collagen genes.";
RL Genomics 4:438-441(1989).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP VARIANT SER-1074.
RX MEDLINE=90036909; PubMed=2572591;
RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
RA Hollister D.W.;
RT "Glycine to serine substitution in the triple helical domain of pro-
RT alpha 1 (II) collagen results in a lethal perinatal form of short-
RT limbed dwarfism.";
RL J. Biol. Chem. 264:18265-18267(1989).
RN [13]
RP VARIANT SEDC GLY-1095--TYR-1330 DEL.
RX MEDLINE=89266907; PubMed=2543071;
RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
RT "Identification of the molecular defect in a family with
RT spondyloepiphyseal dysplasia.";
RL Science 244:978-980(1989).
RN [14]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE=90370826; PubMed=1975693;
RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
RT "Single base mutation in the type II procollagen gene (COL2A1) as a
RT cause of primary osteoarthritis associated with a mild
RT chondrodysplasia.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
RN [15]
RP VARIANT OI-IV VAL-717.
RX MEDLINE=91291136; PubMed=2064612;
RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
RN [16]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE=91086471; PubMed=1985108;
RA Eyre D.R., Weis M.A., Moskowitz R.W.;
RT "Cartilage expression of a type II collagen mutation in an inherited
RT form of osteoarthritis associated with a mild chondrodysplasia.";
RL J. Clin. Invest. 87:357-361(1991).
RN [17]
RP VARIANT HYPOCHONDROGENESIS GLU-984.
RX MEDLINE=93054548; PubMed=1429602;
RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
RA Cohn D.H., Eyre D.R.;
RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
RT 1(II) chain produces hypochondrogenesis.";
RL J. Biol. Chem. 267:22522-22526(1992).
RN [18]
RP VARIANT HYPOCHONDROGENESIS SER-705.
RX MEDLINE=92262484; PubMed=1374906;
RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
RA Ramirez F., Vitale E., Lee B.;
RT "Characterization of a type II collagen gene (COL2A1) mutation
RT identified in cultured chondrocytes from human hypochondrogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
RN [19]
RP VARIANT WS-II ASP-198.
RX MEDLINE=93304428; PubMed=8317498;
RA Koerkkoe J., Ritvaniemi P., Haataja L., Kaeaeriaeinen H.,
RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT for glycine alpha 1-67 and that causes cataracts and retinal
RT detachment: evidence for molecular heterogeneity in the Wagner
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL Am. J. Hum. Genet. 53:55-61(1993).
RN [20]
RP VARIANT SEMD CYS-840.
RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
RA Eyre D.R.;
RT "A dominant mutation in the type II collagen gene (COL2A1) produces
RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
RL Am. J. Hum. Genet. 53:A209-A209(1993).
RN [21]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE=93282819; PubMed=8507190;
RA Holderbaum D., Malemud C.J., Moskowitz R.W., Haqqi T.M.;
RT "Human cartilage from late stage familial osteoarthritis transcribes
RT type II collagen mRNA encoding a cysteine in position 519.";
RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
RN [22]
RP VARIANT SEMD ARG-285.
RX MEDLINE=93252400; PubMed=8486375;
RA Vikkula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
RA Peltonen L.;
RT "A mutation in the amino-terminal end of the triple helix of type II
RT collagen causing severe osteochondrodysplasia.";
RL Genomics 16:282-285(1993).
RN [23]
RP VARIANT SEDC CYS-206.
RX MEDLINE=94063862; PubMed=8244341;
RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
RT family with an Arg75-->Cys mutation in the procollagen type II gene
RT (COL2A1).";
RL Hum. Genet. 92:499-505(1993).
RN [24]
RP VARIANT SEDC CYS-920.
RX MEDLINE=93315508; PubMed=8325895;
RA Chan D., Taylor T.K.F., Cole W.G.;
RT "Characterization of an arginine 789 to cysteine substitution in
RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
RT dysplasia.";
RL J. Biol. Chem. 268:15238-15245(1993).
RN [25]
RP VARIANT SEDC SER-1128.
RX MEDLINE=93140139; PubMed=8423604;

```
RA   Cole W.G., Hall R.K., Rogers J.G.;
RT   "The clinical features of spondyloepiphyseal dysplasia congenita
RT   resulting from the substitution of glycine 997 by serine in the alpha
RT   1(II) chain of type II collagen.";
RL   J. Med. Genet. 30:27-35(1993).

  Query Match            79.5%;  Score 461;  DB 1;  Length 1418;
  Best Local Similarity  76.0%;  Pred. No. 1.3e-20;
  Matches   76;  Conservative    9;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| || |:||:||||||:|||| |||||  |:|| || :|||:|||||||. |  ||
Db    1046 RGDKGEAGEPGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1105

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:|||||||||||||:|: || |||| ||||||||||
Db    1106 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP 1145


RESULT   9
CA12_CHICK
ID   CA12_CHICK     STANDARD;      PRT;   369 AA.
AC   P02460;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   COLLAGEN ALPHA 1(II) CHAIN PRECURSOR (FRAGMENT).
GN   COL2A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-193 FROM N.A.
RX   MEDLINE=85306862; PubMed=3840018;
RA   Deak F., Argraves W.S., Kiss I., Sparks K.J., Goetinck P.F.;
RT   "Primary structure of the telopeptide and a portion of the helical
RT   domain of chicken type II procollagen as determined by DNA sequence
RT   analysis.";
RL   Biochem. J. 229:189-196(1985).
RN   [2]
RP   SEQUENCE OF 82-369 FROM N.A.
RX   MEDLINE=84239728; PubMed=6330084;
RA   Sandell L.J., Prentice H.L., Kravis D., Upholt W.B.;
RT   "Structure and sequence of the chicken type II procollagen gene.
RT   Characterization of the region encoding the carboxyl-terminal
RT   telopeptide and propeptide.";
RL   J. Biol. Chem. 259:7826-7834(1984).
RN   [3]
RP   SEQUENCE OF 114-369 FROM N.A.
RA   Ninomiya Y., Showalter A.M., van der Rest M., Seidah N.G.,
RA   Chretien M., Olsen B.R.;
RT   "Structure of the carboxyl propeptide of chicken type II procollagen
RT   determined by DNA and protein sequence analysis.";
RL   Biochemistry 23:617-624(1984).
CC   -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02663; CAA26499.1; -.
DR   EMBL; L00063; AAB59967.1; -.
DR   EMBL; L00061; AAB59967.1; JOINED.
DR   EMBL; L00062; AAB59967.1; JOINED.
DR   PIR; A02860; CGCH6C.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Cartilage; Collagen.
FT   NON_TER       1     1
FT   CHAIN        <1    123       COLLAGEN ALPHA 1(II) CHAIN.
FT   PROPEP      124    369       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       <1     96       TRIPLE-HELICAL REGION.
FT   DOMAIN       97    123       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD    270    270       N-LINKED (GLCNAC. . .).
FT   DISULFID    275    320
SQ   SEQUENCE   369 AA;  38989 MW;  EF5306925B0BA3B0 CRC64;


  Query Match            75.7%;  Score 439;  DB 1;  Length 369;
  Best Local Similarity  75.0%;  Pred. No. 9.6e-20;
  Matches   72;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       5 GETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGKDGLN 64
           ||||| |:||:||||||:|||| |||||  |:|| :| :||:|||||||  |  |||| |
Db       1 GETGEAGERGLKGHRGFTGLQGLPGPPGPSGDQGAAGPAGPSGPRGPPGPVGPSGKDGSN 60

Qy      65 GLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           |:|||||||||||:|:|: || |||| ||||||||||
Db      61 GMPGPIGPPGPRGRSGEPGPAGPPGNPGPPGPPGPP 96


RESULT  10
CA25_HUMAN
ID   CA25_HUMAN     STANDARD;      PRT;  1496 AA.
AC   P05997;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 2(V) CHAIN PRECURSOR.
GN   COL5A2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-463 FROM N.A.
RX   MEDLINE=89123368; PubMed=2914927;
RA   Woodbury D., Benson-Chanda V., Ramirez F.;
RT   "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
RT   to the structural criteria of a fibrillar procollagen molecule.";
RL   J. Biol. Chem. 264:2735-2738(1989).
RN   [2]
RP   SEQUENCE OF 398-1496 FROM N.A.
RX   MEDLINE=87146331; PubMed=3029669;
RA   Weil D., Bernard M.P., Gargano S., Ramirez F.;
RT   "The pro alpha 2(V) collagen gene is evolutionarily related to the
RT   major fibrillar-forming collagens.";
RL   Nucleic Acids Res. 15:181-198(1987).
RN   [3]
RP   SEQUENCE OF 1227-1496 FROM N.A.
RX   MEDLINE=85289337; PubMed=2411731;
RA   Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
RT   "Complete primary structure of the human alpha 2 type V procollagen
RT   COOH-terminal propeptide.";
RL   J. Biol. Chem. 260:11216-11222(1985).
RN   [4]
RP   SEQUENCE OF 1449-1496 FROM N.A.
RX   MEDLINE=89138450; PubMed=3224983;
RA   Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
RA   Ramirez F.;
```

```
RT   "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
RT   located on the long arm of human chromosome 2.";
RL   Genomics 3:275-277(1988).
RN   [5]
RP   SEQUENCE OF 208-227.
RC   TISSUE=Placenta;
RX   MEDLINE=92239022; PubMed=1571108;
RA   Mann K.;
RT   "Isolation of the alpha 3-chain of human type V collagen and
RT   characterization by partial sequencing.";
RL   Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN   [6]
RP   SEQUENCE OF 288-297 AND 606-617.
RC   TISSUE=Bone;
RX   MEDLINE=94237164; PubMed=8181482;
RA   Moradi-Ameli M., Rousseau J.C., Kleman J.P., Champliaud M.F.,
RA   Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT   "Diversity in the processing events at the N-terminus of type-V
RT   collagen.";
RL   Eur. J. Biochem. 221:987-995(1994).
CC   -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE
CC       COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
CC       TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
CC   -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN
CC       MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
CC       ONE ALPHA 3(V) CHAINS IN PLACENTA.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04478; AAA51859.1; -.
DR   EMBL; X04758; CAA28454.1; -.
DR   EMBL; M11718; AAA52058.1; -.
DR   PIR; A25374; A25374.
DR   PIR; A25874; A25874.
DR   PIR; A30017; A30017.
DR   PIR; A31427; A31427.
DR   MIM; 120190; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     26
FT   CHAIN        27   1226       COLLAGEN ALPHA 2(V) CHAIN.
FT   PROPEP     1227   1496       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       39     97       VWFC.
FT   MOD_RES     290    290       HYDROXYLATION.
FT   MOD_RES     293    293       HYDROXYLATION.
FT   MOD_RES     296    296       HYDROXYLATION.
FT   MOD_RES     608    608       HYDROXYLATION.
FT   MOD_RES     614    614       HYDROXYLATION.
FT   CONFLICT    292    292       A -> P (IN REF. 6).
FT   CONFLICT   1418   1418       K -> T (IN REF. 3).
FT   CONFLICT   1438   1438       F -> S (IN REF. 3).
FT   CONFLICT   1460   1460       E -> Q (IN REF. 4).
FT   CONFLICT   1496   1496       V -> A (IN REF. 4).
SQ   SEQUENCE   1496 AA;  144720 MW;  82827C17A8644F5A CRC64;


  Query Match             69.7%;  Score 404;  DB 1;  Length 1496;
  Best Local Similarity   70.0%;  Pred. No. 2.8e-17;
  Matches   70;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||: |::|||| |||||||:|||| |||||  |||| :|  || |||||||  |  ||
Db    1124 RGDKGDHGDRGDRGQKGHRGFTGLQGLPGPPGPNGEQGSAGIPGPFGPRGPPGPVGPSGK 1183

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           :|  |  ||:|||| ||  |:||| |||| ||||||||||
Db    1184 EGNPGPLGPLGPPGVRGSVGEAGPEGPPGEPGPPGPPGPP 1223


RESULT  11
CA13_MOUSE
ID   CA13_MOUSE     STANDARD;      PRT;  1464 AA.
AC   P08121; Q61429; Q9CRN7;
DT   01-AUG-1988 (Rel. 08, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(III) CHAIN PRECURSOR.
GN   COL3A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL X DBA; TISSUE=Embryo;
RX   MEDLINE=95011609; PubMed=7926795;
RA   Toman D., de Crombrugghe B.;
RT   "The mouse type-III procollagen-encoding gene: genomic cloning and
RT   complete DNA sequence.";
RL   Gene 147:161-168(1994).
RN   [2]
RP   SEQUENCE OF 1-488 FROM N.A.
RX   MEDLINE=88167858; PubMed=3443309;
RA   Wood L., Theriault N., Vogeli G.;
RT   "Complete nucleotide sequence of the N-terminal domains of the murine
RT   alpha-1 type-III collagen chain.";
RL   Gene 61:225-230(1987).
RN   [3]
RP   SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE=85131189; PubMed=3972847;
RA   Liau G., Mudryj M., de Crombrugghe B.;
RT   "Identification of the promoter and first exon of the mouse alpha 1
RT   (III) collagen gene.";
RL   J. Biol. Chem. 260:3773-3777(1985).
RN   [4]
RP   SEQUENCE OF 810-1464 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
```

```
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [5]
RP   SEQUENCE OF 1442-1464 FROM N.A.
RC   STRAIN=C57BL;
RX   MEDLINE=91274355; PubMed=2054384;
RA   Metsaeranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Specific hybridization probes for mouse type I, II, III and IX
RT   collagen mRNAs.";
RL   Biochim. Biophys. Acta 1089:241-243(1991).
CC   -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC       ALONG WITH TYPE I COLLAGEN.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC       LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC       ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC       O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52046; CAA36279.1; -.
DR   EMBL; M18933; AAA37338.1; -.
DR   EMBL; K03037; -; NOT_ANNOTATED_CDS.
DR   EMBL; AK019448; BAB31724.1; -.
DR   EMBL; X57983; CAA41048.1; -.
DR   PIR; A22287; A22287.
DR   PIR; A27353; A27353.
DR   PIR; S16373; S16373.
DR   MGD; MGI:88453; Col3a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     23       BY SIMILARITY.
FT   PROPEP       24    154       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN       155   1203       COLLAGEN ALPHA 1(III) CHAIN.
FT   PROPEP     1204   1464       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       31     90       VWFC.
FT   DOMAIN      155    169       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN      170   1195       TRIPLE-HELICAL REGION.
FT   DOMAIN     1196   1464       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD    262    262       O-LINKED (GAL. . .) (BY SIMILARITY).
FT   MOD_RES     262    262       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     283    283       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     859    859       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     976    976       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    1093   1093       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    1105   1105       HYDROXYLATION (BY SIMILARITY).
FT   DISULFID   1195   1195       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1196   1196       INTERCHAIN (BY SIMILARITY).
SQ   SEQUENCE   1464 AA;  138944 MW;  2104EC27A886090B CRC64;


  Query Match             69.0%;  Score 400;  DB 1;  Length 1464;
  Best Local Similarity   70.0%;  Pred. No. 4.8e-17;
  Matches   70;  Conservative    5;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||||||:|  |||||||| |  |||| ||: | ||  |: |||||||| |  | |||
Db    1090 RGDKGETGERGSNGIKGHRGFPGNPGPPGSPGAAGHQGAIGSPGPAGPRGPVGPHGPPGK 1149

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || :| ||||||||||  |: |  | || || |||||||
Db    1150 DGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPP 1189


RESULT  12
CA13_RAT
ID   CA13_RAT       STANDARD;      PRT;   636 AA.
AC   P13941; O70604;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(III) CHAIN (FRAGMENT).
GN   COL3A1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94114571; PubMed=8286415;
RA   Glumoff V., Maekelae J.K., Vuorio E.;
RT   "Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different
RT   expression patterns of type I and type III collagen and fibronectin
RT   genes in experimental granulation tissue.";
RL   Biochim. Biophys. Acta 1217:41-48(1994).
RN   [2]
RP   SEQUENCE OF 73-636 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Fibroblast;
RA   Wurtz T., Ellerstroem C., Lundmark C., Christersson C.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 308-482 FROM N.A.
RX   MEDLINE=88296083; PubMed=2456904;
RA   Frankel F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lyttle C.R.,
RA   Komm B., Mohn K.;
RT   "Regulation of alpha 2(I), alpha 1(III), and alpha 2(V) collagen
RT   mRNAs by estradiol in the immature rat uterus.";
RL   DNA 7:347-354(1988).
CC   -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC       ALONG WITH TYPE I COLLAGEN.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC       LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC       ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X70369; CAA49832.1; -.
DR   EMBL; AJ005395; CAA06510.1; -.
DR   EMBL; M21354; AAA40942.1; -.
DR   PIR; A29905; A29905.
DR   PIR; S41067; S41067.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 6.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Collagen; Glycoprotein.
```

```
FT   NON_TER       1      1
FT   CHAIN        <1    375       COLLAGEN ALPHA 1(III) CHAIN.
FT   PROPEP      376    636       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       <1    368       TRIPLE-HELICAL REGION.
FT   DOMAIN      369    636       NONHELICAL REGION (C-TERMINAL).
FT   DISULFID    368    368       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    369    369       INTERCHAIN (BY SIMILARITY).
FT   CONFLICT    340    340       N -> D (IN REF. 2).
FT   CONFLICT    429    429       A -> G (IN REF. 2).
SQ   SEQUENCE   636 AA;  62332 MW;  61A48159F01D01EE CRC64;


  Query Match             68.8%;  Score 399;  DB 1;  Length 636;
  Best Local Similarity   70.0%;  Pred. No. 3e-17;
  Matches   70;  Conservative    5;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||||||:|  ||||||||  |  |||| ||: | ||  |: ||||||||  |  | |||
Db    263 RGDKGETGERGSNGIKGHRGFPGNPGPPGSPGAAGHQGAVGSPGPAGPRGPVGPHGPPGK 322

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPGPP 100
          || :| ||||||||||  |: |  | || || |||||||
Db    323 DGSSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPP 362


RESULT  13
CA21_CHICK
ID   CA21_CHICK     STANDARD;      PRT;  1362 AA.
AC   P02467; Q90795; Q90797; P87492; Q90758; Q92014; P87491; Q90792;
DT   21-JUL-1986 (Rel. 01, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 2(I) CHAIN PRECURSOR (FRAGMENTS).
GN   COL1A2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-245; 262-448 AND 466-1362 FROM N.A.
RX   MEDLINE=86185168; PubMed=3868961;
RA   Boedtker H., Finer M., Aho S.;
RT   "The structure of the chicken alpha 2 collagen gene.";
RL   Ann. N.Y. Acad. Sci. 460:85-116(1985).
RN   [2]
RP   SEQUENCE OF 1-89 FROM N.A.
RX   MEDLINE=83246518; PubMed=6135195;
RA   Tate V.E., Finer M.H., Boedtker H., Doty P.;
RT   "Chick pro alpha 2 (I) collagen gene: exon location and coding
RT   potential for the prepropeptide.";
RL   Nucleic Acids Res. 11:91-104(1983).
RN   [3]
RP   SEQUENCE OF 1-14 FROM N.A.
RX   MEDLINE=82060240; PubMed=6946474;
RA   Vogeli G., Ohkubo H., Sobel M.E., Yamada Y., Pastan I.,
RA   de Crombrugghe B.;
RT   "Structure of the promoter for chicken alpha 2 type I collagen gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).
RN   [4]
RP   SEQUENCE OF 1-33 FROM N.A.
RX   MEDLINE=84297217; PubMed=6473103;
RA   Aho S., Tate V.E., Boedtker H.;
RT   "Location of the 11 bp exon in the chicken pro alpha 2(I) collagen
RT   gene.";
RL   Nucleic Acids Res. 12:6117-6125(1984).
RN   [5]
RP   SEQUENCE OF 1-79 FROM N.A.
RX   MEDLINE=88056316; PubMed=3678834;
RA   Finer M.H., Boedtker H., Doty P.;
RT   "Construction and characterization of cDNA clones encoding the 5' end
RT   of the chicken pro alpha 1(I) collagen mRNA.";
RL   Gene 56:71-78(1987).
RN   [6]
RP   SEQUENCE OF 78-92.
RC   TISSUE=Skin;
RX   MEDLINE=71115216; PubMed=5544653;
RA   Highberger J.H., Kang A.H., Gross J.;
RT   "Comparative studies on the amino acid sequence of the alpha 2-CB2
RT   peptides from chick and rat skin collagens.";
RL   Biochemistry 10:610-616(1971).
RN   [7]
RP   SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.
RX   MEDLINE=82058081; PubMed=6272119;
RA   Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.;
RT   "Structure of the pro alpha 2 (I) collagen gene.";
RL   Nature 294:129-135(1981).
RN   [8]
RP   SEQUENCE OF 78-92.
RC   TISSUE=Skin;
RX   MEDLINE=70131186; PubMed=4313735;
RA   Kang A.H., Gross J.;
RT   "Amino acid sequence of cyanogen bromide peptides from the amino-
RT   terminal region of chick skicollagen.";
RL   Biochemistry 9:796-804(1970).
RN   [9]
RP   SEQUENCE OF 78-92 AND 415-448.
RC   TISSUE=Skin;
RX   MEDLINE=69285369; PubMed=5809220;
RA   Kang A.H., Igarashi S., Gross J.;
RT   "Characterization of the cyanogen bromide peptides from the alpha-2
RT   chain of chick skin collagen.";
RL   Biochemistry 8:3200-3204(1969).
RN   [10]
RP   SEQUENCE OF 78-92 AND 415-448.
RC   TISSUE=Bone;
RX   MEDLINE=69206882; PubMed=5785233;
RA   Lane J.M., Miller E.J.;
RT   "Isolation and characterization of the peptides derived from the
RT   alpha 2 chain of chick bone collagen after cyanogen bromide
RT   cleavage.";
RL   Biochemistry 8:2134-2139(1969).
RN   [11]
RP   SEQUENCE OF 566-587 FROM N.A.
RX   MEDLINE=79074829; PubMed=364479;
RA   Lehrach H., Frischauf A.-M., Hanahan D., Wozney J., Fuller F.,
RA   Crkvenjakov R., Boedtker H., Doty P.;
RT   "Construction and characterization of a 2.5-kilobase procollagen
RT   clone.";
RL   Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421(1978).
RN   [12]
RP   SEQUENCE OF 902-1362 FROM N.A.
RX   MEDLINE=81160715; PubMed=6927845;
RA   Fuller F., Boedtker H.;
RT   "Sequence determination and analysis of the 3' region of chicken pro-
RT   alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT   including the carboxy-terminal propeptide sequences.";
RL   Biochemistry 20:996-1006(1981).
RN   [13]
RP   SEQUENCE OF 998-1169 AND 1234-1362 FROM N.A.
RX   MEDLINE=81264246; PubMed=6267043;
RA   Dickson L.A., Ninomiya Y., Bernard M.P., Pesciotta D.M., Parsons J.,
RA   Green G., Eikenberry E.F., de Crombrugghe B., Vogeli G., Pastan I.,
RA   Fietzek P.P., Olsen B.R.;
RT   "The exon/intron structure of the 3'-region of the pro alpha 2(I)
RT   collagen gene.";
RL   J. Biol. Chem. 256:8407-8415(1981).
RN   [14]
RP   SEQUENCE OF 932-954 AND 968-980 FROM N.A.
RX   MEDLINE=81064671; PubMed=6159982;
RA   Avvedimento V.E., Vogeli G., Yamada Y., Maizel J.V. Jr., Pastan I.,
RA   de Crombrugghe B.;
RT   "Correlation between splicing sites within an intron and their
RT   sequence complementarity with U1 RNA.";
RL   Cell 21:689-696(1980).
RN   [15]
```

```
RP   SEQUENCE OF 126-161; 467-517 AND 926-954 FROM N.A.
RX   MEDLINE=81112157; PubMed=7460017;
RA   Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,
RA   Irani M., Pastan I., de Crombrugghe B.;
RT   "The collagen gene: evidence for its evolutionary assembly by
RT   amplification of a DNA segment containing an exon of 54 bp.";
RL   Cell 22:887-892(1980).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25963; AAA69960.1; -.
DR   EMBL; M25956; AAA69960.1; JOINED.
DR   EMBL; M25959; AAA69960.1; JOINED.
DR   EMBL; M25961; AAA69960.1; JOINED.
DR   EMBL; M25962; AAA69960.1; JOINED.
DR   EMBL; M25965; AAA69961.1; -.
DR   EMBL; M25964; AAA69961.1; JOINED.
DR   EMBL; M25984; AAA69962.1; -.
DR   EMBL; M25957; AAA69962.1; JOINED.
DR   EMBL; M25966; AAA69962.1; JOINED.
DR   EMBL; M25967; AAA69962.1; JOINED.
DR   EMBL; M25969; AAA69962.1; JOINED.
DR   EMBL; M25970; AAA69962.1; JOINED.
DR   EMBL; M25971; AAA69962.1; JOINED.
DR   EMBL; M25972; AAA69962.1; JOINED.
DR   EMBL; M25973; AAA69962.1; JOINED.
DR   EMBL; M25974; AAA69962.1; JOINED.
DR   EMBL; M25976; AAA69962.1; JOINED.
DR   EMBL; M25977; AAA69962.1; JOINED.
DR   EMBL; M25978; AAA69962.1; JOINED.
DR   EMBL; M25979; AAA69962.1; JOINED.
DR   EMBL; M25980; AAA69962.1; JOINED.
DR   EMBL; M25981; AAA69962.1; JOINED.
DR   EMBL; M25982; AAA69962.1; JOINED.
DR   EMBL; M25983; AAA69962.1; JOINED.
DR   EMBL; J00826; AAA51611.1; -.
DR   EMBL; J00821; AAA51611.1; JOINED.
DR   EMBL; K00792; AAA51611.1; JOINED.
DR   EMBL; J00830; AAA51613.1; -.
DR   EMBL; J00829; AAA51613.1; JOINED.
DR   EMBL; J00837; AAA51614.1; -.
DR   EMBL; J00812; AAA51615.1; -.
DR   EMBL; J00811; AAA51615.1; JOINED.
DR   EMBL; J00814; AAA51615.1; JOINED.
DR   EMBL; J00815; AAA51615.1; JOINED.
DR   EMBL; X02657; CAA26493.1; -.
DR   EMBL; K00794; -; NOT_ANNOTATED_CDS.
DR   EMBL; V00390; CAA23688.1; -.
DR   EMBL; X00760; CAA25330.1; -.
DR   EMBL; M17608; AAA48673.1; -.
DR   EMBL; M10581; AAA48637.1; -.
DR   EMBL; M10540; AAA48638.1; -.
DR   EMBL; J00828; AAA51612.1; -.
DR   EMBL; J00827; AAA51612.1; JOINED.
DR   EMBL; J00831; -; NOT_ANNOTATED_CDS.
DR   EMBL; J00832; -; NOT_ANNOTATED_CDS.
DR   EMBL; J00833; -; NOT_ANNOTATED_CDS.
DR   EMBL; J00822; -; NOT_ANNOTATED_CDS.
DR   PIR; A02868; CGCH2S.
DR   PIR; S07354; S07354.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001670; Fe-ADH.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01391; Collagen; 16.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.

  Query Match             68.1%;  Score 395;  DB 1;  Length 1362;
  Best Local Similarity   68.7%;  Pred. No. 8.8e-17;
  Matches   68;  Conservative    9;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||:||| |::| ||: | :| :|||| ||  |  |:||| | :|||||||||| :| |||
Db    1003 RGEKGEPGDKGHRGLPGLKGHNGLQGLPGLAGQHGDQGPPGNNGPAGPRGPPGPSGPPGK 1062

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGP 99
           || |||||||||| | ||  |  || |||||||||||||||
Db    1063 DGRNGLPGPIGPAGVRGSHGSQGPAGPPGPPGPPGPPGP 1101


RESULT  14
CA13_HUMAN
ID   CA13_HUMAN     STANDARD;      PRT;  1466 AA.
AC   P02461; Q15112;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(III) CHAIN PRECURSOR.
GN   COL3A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89350838; PubMed=2764886;
RA   Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA   Prockop D.J.;
RT   "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT   chain of human type III procollagen. Differences in protein structure
RT   from type I procollagen and conservation of codon preferences.";
RL   Biochem. J. 260:509-516(1989).
RN   [2]
RP   SEQUENCE OF 149-1225 FROM N.A.
RX   MEDLINE=89386015; PubMed=2780304;
RA   Janeczko R.A., Ramirez F.;
RT   "Nucleotide and amino acid sequences of the entire human alpha 1
RT   (III) collagen.";
RL   Nucleic Acids Res. 17:6742-6742(1989).
RN   [3]
RP   SEQUENCE OF 168-398.
RX   MEDLINE=77134724; PubMed=557335;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of cyanogen
RT   bromide peptides from the amino-terminal segment of type III collagen
RT   of human liver.";
RL   Biochemistry 16:1158-1164(1977).
RN   [4]
RP   REVISIONS.
RA   Seyer J.M.;
RL   Submitted (DEC-1977) to the PIR data bank.
RN   [5]
RP   SEQUENCE OF 399-727.
RX   MEDLINE=79000343; PubMed=687591;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of five
RT   consecutive CNBr peptides from type III collagen of human liver.";
RL   Biochemistry 17:3404-3411(1978).
RN   [6]
RP   SEQUENCE OF 728-964.
```

RX MEDLINE=80198282; PubMed=6246925;
RA Seyer J.M., Mainardi C., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1
RT (III)-CB5 from type III collagen of human liver.";
RL Biochemistry 19:1583-1589(1980).
RN [7]
RP SEQUENCE OF 950-1466 FROM N.A.
RX MEDLINE=88189827; PubMed=3357782;
RA Mankoo B.S., Dalgleish R.;
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL Nucleic Acids Res. 16:2337-2337(1988).
RN [8]
RP REVISION TO 1184.
RX MEDLINE=89098346; PubMed=3211760;
RA Molyneux K., Dalgleish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:11833-11833(1988).
RN [9]
RP SEQUENCE OF 1065-1466 FROM N.A.
RX MEDLINE=85087944; PubMed=6096827;
RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA Rosenbloom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III
RT procollagen.";
RL Nucleic Acids Res. 12:9383-9394(1984).
RN [10]
RP SEQUENCE OF 965-1200.
RX MEDLINE=81208139; PubMed=7016180;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha
RT 1(III)-CB9 from type III collagen of human liver.";
RL Biochemistry 20:2621-2627(1981).
RN [11]
RP SEQUENCE OF 1176-1466 FROM N.A.
RX MEDLINE=85157600; PubMed=2579949;
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT (III) collagen. Partial characterization of the 3' end region of the
RT gene.";
RL J. Biol. Chem. 260:4357-4363(1985).
RN [12]
RP SEQUENCE OF 1161-1200 FROM N.A.
RX MEDLINE=86187804; PubMed=3754462;
RA Miskulin M., Dalgleish R., Kluve-Beckerman B., Rennard S.I.,
RA Tolstoshev P., Brantly M., Crystal R.G.;
RT "Human type III collagen gene expression is coordinately modulated
RT with the type I collagen genes during fibroblast growth.";
RL Biochemistry 25:1408-1413(1986).
RN [13]
RP SEQUENCE OF 1-170 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88303360; PubMed=3405773;
RA Toman D., Ricca G., de Crombrugghe B.;
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
RT of human prepro alpha 1(III) collagen.";
RL Nucleic Acids Res. 16:7201-7201(1988).
RN [14]
RP SEQUENCE OF 1-176 FROM N.A.
RX MEDLINE=89378752; PubMed=2777083;
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT "Cloning and analysis of the 5' portion of the human type-III
RT procollagen gene (COL3A1).";
RL Gene 78:255-265(1989).
RN [15]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [16]
RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
RX MEDLINE=93293988; PubMed=8514866;
RA Tromp G., Wu Y., Prockop D.J., Madhatheri S.L., Kleinert C.,
RA Earley J.J., Zhuang J., Noerrgaard O., Darling R.C., Abbott W.M.,
RA Cole C.W., Jaakkola P., Ryynaenen M., Pearce W.H., Yao J.S.T.,
RA Majamaa K., Smullens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
RT in the triple-helical domain of type III procollagen are an
RT infrequent cause of aortic aneurysms.";
RL J. Clin. Invest. 91:2539-2545(1993).
RN [17]
RP VARIANT THR-698.
RX MEDLINE=91045136; PubMed=2235526;
RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
RA Wu Y., Ganguly A., Prockop D.J.;
RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
RL Nucleic Acids Res. 18:6180-6180(1990).
RN [18]
RP VARIANT AORTIC ANEURYSM ARG-786.
RX MEDLINE=91056145; PubMed=2243125;
RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
RT "A mutation in the gene for type III procollagen (COL3A1) in a family
RT with aortic aneurysms.";
RL J. Clin. Invest. 86:1465-1473(1990).
RN [19]
RP VARIANT EDS-IV ARG-828.
RX MEDLINE=94016385; PubMed=8411057;
RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
RT "The substitution of glycine 661 by arginine in type III collagen
RT produces mutant molecules with different thermal stabilities and
RT causes Ehlers-Danlos syndrome type IV.";
RL J. Med. Genet. 30:690-693(1993).
RN [20]
RP VARIANT EDS-IV SER-957.
RX MEDLINE=89109135; PubMed=2492273;
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
RT "A single base mutation that substitutes serine for glycine 790 of
RT the alpha 1 (III) chain of type III procollagen exposes an arginine
RT and causes Ehlers-Danlos syndrome IV.";
RL J. Biol. Chem. 264:1349-1352(1989).
RN [21]
RP VARIANT EDS-IV VAL-960.
RX MEDLINE=95268429; PubMed=7749417;
RA Tromp G., de Paepe A., Nuytinck L., Madhatheri S.L., Kuivaniemi H.;
RT "Substitution of valine for glycine 793 in type III procollagen in
RT Ehlers-Danlos syndrome type IV.";
RL Hum. Mutat. 5:179-181(1995).
RN [22]
RP VARIANT EDS-IV GLU-1014.
RX MEDLINE=92316511; PubMed=1352273;
RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
RA Pope F.M.;
RT "A single base mutation in the gene for type III collagen (COL3A1)
RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT syndrome type IV. An unaffected family member is mosaic for the
RT mutation.";
RL Hum. Genet. 89:414-418(1992).
RN [23]
RP VARIANT EDS-IV ASP-1050.
RX MEDLINE=90037070; PubMed=2808425;
RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
RT "Single base mutation in the type III procollagen gene that converts
RT the codon for glycine 883 to aspartate in a mild variant of
RT Ehlers-Danlos syndrome IV.";
RL J. Biol. Chem. 264:19313-19317(1989).
RN [24]
RP VARIANT EDS-IV VAL-1077.
RX MEDLINE=91374480; PubMed=1895316;
RA Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,
RA Pope F.M.;
RT "Characterisation of a glycine to valine substitution at amino acid
RT position 910 of the triple helical region of type III collagen in a
RT patient with Ehlers-Danlos syndrome type IV.";

```
RL   J. Med. Genet. 28:458-463(1991).
RN   [25]
RP   VARIANT EDS-IV GLU-1173.
RX   MEDLINE=93022543; PubMed=1357232;
RA   Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;

  Query Match             67.6%;  Score 392;  DB 1;  Length 1466;
  Best Local Similarity   69.0%;  Pred. No. 1.4e-16;
  Matches   69;  Conservative    6;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||||||:|  ||||||||  |  | ||  ||  |:||  |: ||||||||  | :| |||
Db    1091 RGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGPAGPRGPVGPSGPPGK 1150

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || :| ||||||||||  |: |  | || || |||||||
Db    1151 DGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPP 1190


RESULT  15
CA13_BOVIN
ID   CA13_BOVIN     STANDARD;      PRT;  1049 AA.
AC   P04258;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(III) CHAIN.
GN   COL3A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-242.
RX   MEDLINE=80026026; PubMed=488906;
RA   Fietzek P.P., Allmann H., Rauterberg J., Henkel W., Wachter E.,
RA   Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. I. The amino
RT   acid sequence of the amino terminal region of the alpha 1(III) chain
RT   (positions 1-222).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
RN   [2]
RP   SEQUENCE OF 243-422.
RX   MEDLINE=80026027; PubMed=488907;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. II. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
RT   (positions 223-402).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
RN   [3]
RP   SEQUENCE OF 423-571.
RX   MEDLINE=80026028; PubMed=488908;
RA   Bentz H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. III. The
RT   amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT   (positions 403-551).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
RN   [4]
RP   SEQUENCE OF 572-808.
RX   MEDLINE=80026029; PubMed=488909;
RA   Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. IV. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT   (positions 552-788).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
RN   [5]
RP   SEQUENCE OF 809-947.
RX   MEDLINE=80026030; PubMed=488910;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. V. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT   (position 789-927).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
RN   [6]
RP   SEQUENCE OF 948-1049.
RX   MEDLINE=80026031; PubMed=488911;
RA   Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. VI. The amino
RT   acid sequence of the carboxyterminal cyanogen bromide peptide alpha
RT   1(III)CB9B (positions 928-1028).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
CC   -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC       ALONG WITH TYPE I COLLAGEN.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC       LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC       ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
DR   PIR; A02862; CGBO7S.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 17.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen.
FT   DOMAIN        1     14       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN       15   1040       TRIPLE-HELICAL REGION.
FT   DOMAIN     1041   1049       NONHELICAL REGION (C-TERMINAL).
FT   MOD_RES      95     95       HYDROXYLATION.
FT   MOD_RES     107    107       HYDROXYLATION.
FT   MOD_RES     119    119       HYDROXYLATION.
FT   MOD_RES     938    938       HYDROXYLATION.
FT   MOD_RES     950    950       HYDROXYLATION.
FT   CARBOHYD    107    107       O-LINKED (GAL. . .).
FT   CARBOHYD    950    950       O-LINKED (GAL. . .).
FT   DISULFID   1040   1040       INTERCHAIN.
FT   DISULFID   1041   1041       INTERCHAIN.
SQ   SEQUENCE   1049 AA;  93651 MW;  8EEC33D1C66EC9A3 CRC64;


  Query Match             67.2%;  Score 390;  DB 1;  Length 1049;
  Best Local Similarity   69.0%;  Pred. No. 1.4e-16;
  Matches   69;  Conservative    5;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||||||:|  ||||||||  |  | ||  ||  | ||  |: ||||||||  | :| |||
Db     935 RGDKGETGERGAMGIKGHRGFPGNPGAPGSPGPAGHQGAVGSPGPAGPRGPVGPSGPPGK 994

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || :| ||||||||||  |: |  | || || |||||||
Db     995 DGASGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPP 1034


Search completed: January 28, 2002, 07:48:33
Job time: 98 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:48:11 ; Search time 37.99 Seconds
(without alignments)
385.029 Million cell updates/sec

Title: US-09-710-239-29
Perfect score: 580
Sequence: 1 RGDKGETGEQGDRGIKGHRG.....DAGPVGPPGPPGPPGPPGPP 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	580	100.0	1461	4	O76045	O76045 homo sapien
2	562	96.9	589	11	Q99LL6	Q99LL6 mus musculu
3	562	96.9	1453	11	Q63079	Q63079 rattus norv
4	507	87.4	1450	13	Q9YIB4	Q9yib4 cynops pyrr
5	503	86.7	1445	13	O93251	O93251 rana catesb
6	480.5	82.8	1447	13	Q9IB91	Q9ib91 xenopus lae
7	480	82.8	809	13	O93485	O93485 oncorhynchu
8	472	81.4	1442	11	Q62031	Q62031 mus musculu
9	472	81.4	1442	11	Q62033	Q62033 mus musculu
10	472	81.4	1459	11	Q62032	Q62032 mus musculu
11	462	79.7	1418	13	Q9W7R9	Q9w7r9 cynops pyrr
12	461	79.5	207	4	Q14044	Q14044 homo sapien
13	461	79.5	1160	4	Q14046	Q14046 homo sapien
14	461	79.5	1418	6	Q28396	Q28396 equus cabal
15	461	79.5	1487	4	Q14047	Q14047 homo sapien
16	461	79.5	1487	6	O77753	O77753 canis famil
17	460	79.3	1419	11	Q63123	Q63123 rattus norv
18	451	77.8	1486	13	Q91717	Q91717 xenopus lae
19	449	77.4	1491	13	Q91718	Q91718 xenopus lae
20	447	77.1	464	13	Q90412	Q90412 brachydanio
21	416	71.7	678	13	O93486	O93486 oncorhynchu
22	415	71.6	469	11	O70598	O70598 rattus norv
23	410	70.7	1497	11	Q61431	Q61431 mus musculu
24	404	69.7	225	4	P78440	P78440 homo sapien
25	400	69.0	655	11	Q9CRN7	Q9crn7 mus musculu
26	397	68.4	564	11	O70604	O70604 rattus norv
27	392	67.6	132	4	P78429	P78429 homo sapien
28	373	64.3	1355	13	O42350	O42350 rana catesb
29	349	60.2	526	6	Q28668	Q28668 oryctolagus
30	349	60.2	1372	11	Q9R1E8	Q9r1e8 rattus norv
31	348	60.0	1366	4	Q9UPH0	Q9uph0 homo sapien
32	346	59.7	1186	4	Q9UEB6	Q9ueb6 homo sapien
33	346	59.7	1366	4	Q15177	Q15177 homo sapien
34	337	58.1	940	13	O93484	O93484 oncorhynchu
35	318	54.8	215	13	O93392	O93392 coturnix co
36	312.5	53.9	1835	13	Q9IAU4	Q9iau4 gallus gall
37	310.5	53.5	1840	11	Q9JI03	Q9ji03 rattus norv
38	309	53.3	1378	5	O97405	O97405 haliotis di
39	308.5	53.2	1838	11	O88207	O88207 mus musculu
40	308.5	53.2	1840	11	Q60467	Q60467 cricetulus
41	307.5	53.0	1838	4	Q15094	Q15094 homo sapien
42	306.5	52.8	1146	13	Q90584	Q90584 gallus gall
43	306	52.8	890	5	O77087	O77087 alvinella p
44	304.5	52.5	230	11	Q9R149	Q9r149 cavia porce
45	302	52.1	501	5	Q9NDD6	Q9ndd6 riftia pach

ALIGNMENTS

RESULT 1
O76045
ID O76045 PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PRO ALPHA 1(I) COLLAGEN.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329734; PubMed=2843432;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1).";
RL Gene 67:105-115(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";

```
RL   FEBS Lett. 279:9-13(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92157916; PubMed=1787829;
RA   Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA   Olsen A.S., Prockop D.J.;
RT   "Completion of the last half of the structure of the human gene for
RT   the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL   Matrix 11:375-379(1991).
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Korkko J.M., Earley J.J., Nuytinck L., DePaepe A., Prockop D.J.,
RA   Ala-Kokko L.;
RT   "Analysis of the COL1A1 and COL1A2 Genes by CSGE and DNA Sequencing in
RT   12 Patients with mild OI (Type I). Identification of Common Sequences
RT   for Null Allele Mutations.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF017178; AAB94054.2; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
KW   Collagen.
SQ   SEQUENCE   1461 AA;  138629 MW;  9ACF6DE30EA78E21 CRC64;


  Query Match             100.0%;  Score 580;  DB 4;  Length 1461;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-44;
  Matches  100;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1090 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 1149

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          |||||||||||||||||||||||||||||||||||||||||
Db   1150 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 1189


RESULT   2
Q99LL6
ID   Q99LL6      PRELIMINARY;      PRT;   589 AA.
AC   Q99LL6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UNKNOWN (PROTEIN FOR IMAGE:3586143) (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC003198; AAH03198.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   589 AA;  58805 MW;  81847495E5E05CEF CRC64;


  Query Match             96.9%;  Score 562;  DB 11;  Length 589;
  Best Local Similarity   96.0%;  Pred. No. 3.3e-43;
  Matches   96;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||:|||
Db    218 RGDKGETGEQGDRGIKGHRGFSGLQGPPGSPGSPGEQGPSGASGPAGPRGPPGSAGSPGK 277

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          ||||||||||||||||||||:|| ||||||||||||||||
Db    278 DGLNGLPGPIGPPGPRGRTGDSGPAGPPGPPGPPGPPGPP 317


RESULT   3
Q63079
ID   Q63079      PRELIMINARY;      PRT;  1453 AA.
AC   Q63079;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COLLAGEN ALPHA1 (FRAGMENT).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1092 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=TOOTH;
RA   Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z78279; CAB01633.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   1453 AA;  137886 MW;  E6896BDC19A4A1D8 CRC64;


  Query Match             96.9%;  Score 562;  DB 11;  Length 1453;
  Best Local Similarity   96.0%;  Pred. No. 7.4e-43;
  Matches   96;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||:|||
Db   1082 RGDKGETGEQGDRGIKGHRGFSGLQGPPGSPGSPGEQGPSGASGPAGPRGPPGSAGSPGK 1141

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          ||||||||||||||||||||:|| ||||||||||||||||
Db   1142 DGLNGLPGPIGPPGPRGRTGDSGPAGPPGPPGPPGPPGPP 1181


RESULT   4
Q9YIB4
ID   Q9YIB4      PRELIMINARY;      PRT;  1450 AA.
AC   Q9YIB4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA 1 TYPE I COLLAGEN.
OS   Cynops pyrrhogaster (Japanese common newt).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX   NCBI_TaxID=8330;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=REGENERATE FORELIMBS;
RA   Asahina K., Obara M., Yoshizato K.;
RT   "Cynops pyrrhogaster alpha 1 type I collagen, partial cDNA.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB015438; BAA36973.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 18.
```

```
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; UNKNOWN_1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
KW   Collagen.
SQ   SEQUENCE   1450 AA;  137563 MW;  ABF8A74841B87B7C CRC64;


  Query Match             87.4%;  Score 507;  DB 13;  Length 1450;
  Best Local Similarity   84.0%;  Pred. No. 6.5e-38;
  Matches   84;  Conservative    7;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| |||||:||:||||||:|:||||||||| ||||  | ||||||||||||:|: ||
Db    1079 RGDKGEAGEQGERGMKGHRGFNGMQGPPGPPGSSGEQGAPGPSGPAGPRGPPGSSGSTGK 1138

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||:|||||||||||||| || || ||||||||||||||
Db    1139 DGVNGLPGPIGPPGPRGRNGDVGPAGPPGPPGPPGPPGPP 1178


RESULT   5
O93251
ID   O93251      PRELIMINARY;      PRT;  1445 AA.
AC   O93251;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA 1 TYPE I COLLAGEN.
OS   Rana catesbeiana (Bull frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX   NCBI_TaxID=8400;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Asahina K., Utoh R., Obara M., Yoshizato K.;
RT   "Spatiotemporal expression of bullfrog a1(I) and a2(I) collagen genes
RT   in intestine during metamorphosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB015440; BAA29028.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
SQ   SEQUENCE   1445 AA;  137251 MW;  F59BB550C9873F04 CRC64;


  Query Match             86.7%;  Score 503;  DB 13;  Length 1445;
  Best Local Similarity   85.0%;  Pred. No. 1.5e-37;
  Matches   85;  Conservative    5;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| ||||:||:|||||| | |||| ||  |||||||||||||||||||:|:|||
Db    1078 RGDKGEAGEQGERGMKGHRGFNDLPGPPGAPGHAGEQGPSGASGPAGPRGPPGSSGSPGK 1137

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || |||||||||||||||| || ||||| |||||||||
Db    1138 DGSNGLPGPIGPPGPRGRTGDVGPAGPPGPAGPPGPPGPP 1177


RESULT   6
Q9IB91
ID   Q9IB91      PRELIMINARY;      PRT;  1447 AA.
AC   Q9IB91;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TYPE I COLLAGEN ALPHA 1.
GN   COL1A1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RT   "Expression and characterization of Xenopus type I collagen alpha 1
RT   (COL1A1) during embryonic development.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034701; BAA94972.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Collagen.
SQ   SEQUENCE   1447 AA;  137445 MW;  A4A6DD2B4158E38B CRC64;


  Query Match             82.8%;  Score 480.5;  DB 13;  Length 1447;
  Best Local Similarity   85.7%;  Pred. No. 1.6e-35;
  Matches   84;  Conservative    4;  Mismatches    9;  Indels    1;  Gaps    1;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| ||||:||:|||||||:|  |||||||| ||||||||||||||||||||:| |||
Db    1076 RGDKGEAGEQGERGMKGHRGFNGPSGPPGPPGSSGEQGPSGASGPAGPRGPPGSSGNPGK 1135

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPG 98
           || ||||| |||||||||||| || |||| ||||||||
Db    1136 DGSNGLPGLIGPPGPRGRTGDVGPAGPPG-PGPPGPPG 1172


RESULT   7
O93485
ID   O93485      PRELIMINARY;      PRT;   809 AA.
AC   O93485;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA 1 TYPE I COLLAGEN (FRAGMENT).
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FIBROBLAST;
RA   Saito M., Kunisaki N., Hirono I., Aoki T., Ishida M., Urano N.,
RA   Kimura S.;
RT   "Partial characterization of cDNA clones encoding the three distinct
RT   pro alpha chains of type I collagen from rainbow trout.";
RL   Fisheries Sci. 64:780-786(1998).
DR   EMBL; AB008373; BAA33380.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01391; Collagen; 9.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   NON_TER       1       1
SQ   SEQUENCE   809 AA;  78164 MW;  68C056A7640FCA81 CRC64;


  Query Match             82.8%;  Score 480;  DB 13;  Length 809;
  Best Local Similarity   80.0%;  Pred. No. 1e-35;
```

```
  Matches   80;  Conservative   10;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||:|| |:||:||||||:|:|||||| |  || ||:||||||||||| ||||| ||
Db     438 RGDKGESGEAGERGMKGHRGFTGMQGPPGPSGQSGESGPAGASGPAGPRGPSGSAGAAGK 497

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           ||::||||||||||||||:|: || | ||||||||||||
Db     498 DGMSGLPGPIGPPGPRGRSGEMGPSGTPGPPGPPGPPGPP 537


RESULT   8
Q62031
ID   Q62031      PRELIMINARY;      PRT;  1442 AA.
AC   Q62031;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PRO-ALPHA-1 TYPE II COLLAGEN.
GN   COL2A1 OR PRO-ALPHA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RX   MEDLINE=91358489; PubMed=1885613;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. Biol. Chem. 266:16862-16869(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (OCT-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M65161; AAA68099.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   1442 AA;  137911 MW;  D3A3274493CB821C CRC64;


  Query Match             81.4%;  Score 472;  DB 11;  Length 1442;
  Best Local Similarity   78.0%;  Pred. No. 9.1e-35;
  Matches   78;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||:||||:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db    1070 RGDKGESGEQGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1129

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:||||||||||||:|: |||||||  ||||||||||
Db    1130 DGSNGIPGPIGPPGPRGRSGETGPVGPPGSPGPPGPPGPP 1169


RESULT   9
Q62033
ID   Q62033      PRELIMINARY;      PRT;  1442 AA.
AC   Q62033;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PRO-ALPHA-1 TYPE II COLLAGEN.
GN   COL2A1 OR PRO-ALPHA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (OCT-1991) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M65161; AAA68102.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   1442 AA;  137828 MW;  F0E77C11BCAFA93B CRC64;


  Query Match             81.4%;  Score 472;  DB 11;  Length 1442;
  Best Local Similarity   78.0%;  Pred. No. 9.1e-35;
  Matches   78;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||:||||:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db    1070 RGDKGESGEQGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1129

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:||||||||||||:|: |||||||  ||||||||||
Db    1130 DGSNGIPGPIGPPGPRGRSGETGPVGPPGSPGPPGPPGPP 1169


RESULT  10
Q62032
ID   Q62032      PRELIMINARY;      PRT;  1459 AA.
AC   Q62032;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PRO-ALPHA-1 TYPE II COLLAGEN.
GN   COL2A1 OR PRO-ALPHA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RX   MEDLINE=91358489; PubMed=1885613;
```

```
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. Biol. Chem. 266:16862-16869(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (OCT-1991) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57/BLACK;
RA   Vuorio E.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; M65161; AAA68101.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   1459 AA;  139070 MW;  A09D24BF7357C827 CRC64;


  Query Match             81.4%;  Score 472;  DB 11;  Length 1459;
  Best Local Similarity   78.0%;  Pred. No. 9.2e-35;
  Matches   78;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||:||||:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db    1087 RGDKGESGEQGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1146

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:||||||||||||:|: ||||||| ||||||||||
Db    1147 DGSNGIPGPIGPPGPRGRSGETGPVGPPGSPGPPGPPGPP 1186


RESULT  11
Q9W7R9
ID   Q9W7R9      PRELIMINARY;      PRT;  1418 AA.
AC   Q9W7R9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA1 TYPE II COLLAGEN.
OS   Cynops pyrrhogaster (Japanese common newt).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX   NCBI_TaxID=8330;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Asahina K., Obara M., Yoshizato K.;
RT   "Unique expression of genes of type I and type II collagens of
RT   regenerating newt limb in apical epidermal cap, blastema, muscle and
RT   cartilage.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB022046; BAA82043.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Collagen.
SQ   SEQUENCE   1418 AA;  135066 MW;  C19A6E601A2A717E CRC64;


  Query Match             79.7%;  Score 462;  DB 13;  Length 1418;
  Best Local Similarity   79.0%;  Pred. No. 7.1e-34;
  Matches   79;  Conservative    5;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||:||:||:||||||:|||| |||||: |:|| || |||||||||||  |  ||
Db    1046 RGDKGEAGEAGERGQKGHRGFTGLQGLPGPPGTAGDQGASGPSGPAGPRGPPGPVGPSGK 1105

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || |||||||||||||| |: || |||| ||||||||||
Db    1106 DGSNGLPGPIGPPGPRGRGGETGPAGPPGNPGPPGPPGPP 1145


RESULT  12
Q14044
ID   Q14044      PRELIMINARY;      PRT;  207 AA.
AC   Q14044;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TYPE II COLLAGEN (FRAGMENT).
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91286276; PubMed=1905723;
RA   Chan D., Cole W.G.;
RT   "Low basal transcription of genes for tissue-specific collagens by
RT   fibroblasts and lymphoblastoid cells. Application to the
RT   characterization of a glycine 997 to serine substitution in alpha
RT   1(II) collagen chains of a patient with spondyloepiphyseal
RT   dysplasia.";
RL   J. Biol. Chem. 266:12487-12494(1991).
DR   EMBL; M63281; AAA52038.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01391; Collagen; 2.
DR   Pfam; PF01410; COLFI; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   207 AA;  20145 MW;  8AE2149116B6425E CRC64;


  Query Match             79.5%;  Score 461;  DB 4;  Length 207;
  Best Local Similarity   76.0%;  Pred. No. 1.6e-34;
  Matches   76;  Conservative    9;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           ||||||:||:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db      34 RGDKGEAGEPGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 93

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:||||||||||||:|: || |||| ||||||||||
Db      94 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP 133


RESULT  13
Q14046
ID   Q14046      PRELIMINARY;      PRT;  1160 AA.
AC   Q14046;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COL2A1 ALPHA1 (II) COLLAGEN PRECURSOR (FRAGMENT).
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CARTILAGE;
```

```
RX   MEDLINE=90026318; PubMed=2803268;
RA   Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
RT   "Structure of cDNA clones coding for human type II procollagen. The
RT   alpha 1(II) chain is more similar to the alpha 1(I) chain than two
RT   other alpha chains of fibrillar collagens.";
RL   Biochem. J. 262:521-528(1989).
DR   EMBL; X16711; CAA34683.1; -.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 18.
KW   Signal; Matrix protein.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN       113  >1160       COLLAGEN.
FT   NON_TER    1160   1160
SQ   SEQUENCE   1160 AA;  105630 MW;  A7F0523B856C8639 CRC64;


  Query Match             79.5%;  Score 461;  DB 4;  Length 1160;
  Best Local Similarity   76.0%;  Pred. No. 7.3e-34;
  Matches   76;  Conservative    9;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||| || |:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db   1046 RGDKGEAGEPGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1105

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          || ||:|||||||||||||:|: || |||| ||||||||||
Db   1106 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP 1145


RESULT  14
Q28396
ID   Q28396      PRELIMINARY;      PRT;  1418 AA.
AC   Q28396;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TYPE II COLLAGEN.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Richardson D.W., Dodge G.R.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 18-68 FROM N.A.
RA   MacLeod J.N., Fubini S.L., Gu D.N., Tetreault J.W., Todhunter R.J.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U62528; AAB05773.1; -.
DR   EMBL; AF040638; AAB96768.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
SQ   SEQUENCE   1418 AA;  134343 MW;  115FCD19EB8696A3 CRC64;


  Query Match             79.5%;  Score 461;  DB 6;  Length 1418;
  Best Local Similarity   76.0%;  Pred. No. 8.8e-34;
  Matches   76;  Conservative    9;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
          |||||| || |:||:||||||:|||| |||||  |:|| || :||:|||||||  |  ||
Db   1046 RGDKGEAGEAGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1105

Qy     61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
          || ||:|||||||||||||:|: || |||| ||||||||||
Db   1106 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP 1145


RESULT  15
Q14047
ID   Q14047      PRELIMINARY;      PRT;  1487 AA.
AC   Q14047;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-1 TYPE II COLLAGEN.
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE=85190534; PubMed=3857598;
RA   Cheah K.S., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
RT   "Identification and characterization of the human type II collagen
RT   gene (COL2A1).";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE=90026318; PubMed=2803268;
RA   Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
RT   "Structure of cDNA clones coding for human type II procollagen. The
RT   alpha 1(II) chain is more similar to the alpha 1(I) chain than two
RT   other alpha chains of fibrillar collagens.";
RL   Biochem. J. 262:521-528(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE=91184811; PubMed=2081599;
RA   Ryan M.C., Sieraski M., Sandell L.J.;
RT   "The human type II procollagen gene: identification of an additional
RT   protein-coding domain and location of potential regulatory sequences
RT   in the promoter and first intron.";
RL   Genomics 8:41-48(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE=91153296; PubMed=1999183;
RA   Huang M.C., Seyer J.M., Thompson J.P., Spinella D.G., Cheah K.S.,
RA   Kang A.H.;
RT   "Genomic organization of the human procollagen alpha 1(II) collagen
RT   gene.";
RL   Eur. J. Biochem. 195:593-600(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE=97104294; PubMed=8948452;
RA   Ala-Kokko L., Kvist A.P., Metsaranta M., Kivirikko K.I.,
RA   Crombrugghe B., Prockop D.J., Vuorio E.;
RT   "Conservation of the sizes of 53 introns and over 100 intronic
RT   sequences for the binding of common transcription factors in the human
RT   and mouse genes for type II procollagen (COL2A1).";
RL   Biochem. J. 308:923-929(1995).
DR   EMBL; L10347; AAC41772.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
SQ   SEQUENCE   1487 AA;  141771 MW;  0B7E79D46BDAFA97 CRC64;


  Query Match             79.5%;  Score 461;  DB 4;  Length 1487;
```

Best Local Similarity 76.0%; Pred. No. 9.1e-34;
Matches 76; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

```
Qy       1 RGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGK 60
           |||||| || |:||:||||||:|||| ||||| |:|| || :||:|||||||  |  ||
Db    1115 RGDKGEAGEPGERGLKGHRGFTGLQGLPGPPGPSGDQGASGPAGPSGPRGPPGPVGPSGK 1174

Qy      61 DGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPP 100
           || ||:|||||||||||||:|: || |||| ||||||||||
Db    1175 DGANGIPGPIGPPGPRGRSGETGPAGPPGNPGPPGPPGPP 1214
```

Search completed: January 28, 2002, 07:48:13
Job time: 78 sec